

The persistent cosmic web and its filamentary structure

II: Illustrations

T. Sousbie^{1,2*}, C. Pichon^{2,3}, H. Kawahara⁴

¹*Department of Physics, The University of Tokyo, Tokyo 113-0033, Japan*

²*Institut d'astrophysique de Paris & UPMC (UMR 7095), 98, bis boulevard Arago, 75 014, Paris, France.*

³*Oxford Astrophysics, department of Physics, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK.*

⁴*Department of Physics, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan*

22 September 2010

ABSTRACT

The recently introduced discrete persistent structure extractor (DisPerSE Soubie 2010, paper I) is implemented on realistic 3D cosmological simulations and observed redshift catalogues; it is found that DisPerSE traces very well the observed filaments, walls, and voids seen both in simulations and observations. In either setting, filaments are shown to connect onto halos, outskirt walls, which circumvent voids, as is topologically required by Morse theory. Indeed this algorithm returns the optimal critical set while operating directly on the particles. DisPerSE, as illustrated here, assumes nothing about the geometry of the survey or its homogeneity, and yields a natural (topologically motivated) self-consistent criterion for selecting the significance level of the identified structures. It is shown that this extraction is possible even for very sparsely sampled point processes, as a function of the persistence ratio (a measure of the significance of topological connections between critical points). Hence astrophysicists should be in a position to trace precisely the locus of filaments, walls and voids from such samples and assess the confidence of the post-processed sets as a function of this threshold, which can be expressed relative to the expected amplitude of shot noise. In a cosmic framework, this criterion is shown to level with friend of friend structure finder for the identifications of peaks, while it also identifies the connected filaments and walls, and quantitatively recovers the full set of topological invariants (number of holes, etc...) *directly from the particles*, and at no extra cost as a function of the persistence threshold. This criterion is found to be sufficient even if one particle out of two is noise, when the persistence ratio is set to 3-sigma or more. The algorithm is also implemented on the SDSS catalogue and used to locate interesting configurations of the filamentary structure. In this context we carried the identification of an “optically faint” cluster at the intersection of filaments through the recent observation of its X-ray counterpart by SUZAKU. The corresponding filament catalogue is available online.

Key words: Cosmology: simulations, statistics, observations, Galaxies: formation, dynamics.

1 INTRODUCTION

Over the past decades, numerical simulations (*e.g.* Efsthathiou et al. 1985), and large redshift surveys (*e.g.* de Lapparent et al. 1986) have highlighted the large-scale structure (hereafter LSS) of our Universe, a cosmic web formed by voids, sheets, elongated filaments and clusters at their nodes (Pogosyan et al. 1996). Characterizing quantitatively

these striking features of the observed and modeled universe has proven to be both useful (Sousbie et al. 2008; Gay et al. 2010) but challenging. It is useful because these features reflect the underlying dynamics of structure formation, and are therefore sensitive to the content of the universe (Pogosyan et al. 2009). It is challenging because observations and simulations provide limited and noisy data sets. Recently Soubie (2010, hereafter paper I) presented an algorithm able to estimate the underlying critical sets (walls, filaments, voids) from a given noisy discrete sample

* tsousbie@gmail.com

of the underlying field. Typically, this situation arises in astrophysics when the aim is to recover the topology or the geometry of the underlying density field while only a catalogue of galaxies is available. For instance, in the context of understanding the history of our Milky Way, it is of interest to identify the filaments of the local group. Yet typically in this context, only a limited number of galaxies at somewhat poorly estimated positions are observed. For redshift catalogues involving hundreds of thousands of galaxies, one would also wish to reconstruct the main features of the cosmic web as best as the non-uniform sampling allows. From a theoretical point of view, it might for instance be of interest to compute the cosmic evolution of the filamentary network, as its history constrains the dark energy content of the universe. From an observational point of view, it could also help solving the missing baryon problem (Fukugita et al. 1998) because most of such baryons has been considered to be located along the filamentary structure in the form of diffuse hot gas called Warm/Hot Intergalactic Medium (WHIM; Cen & Ostriker (1999), Aracil et al. (2004)). Identifying the filament from galaxy distributions clearly provides good candidates for searching for the WHIM with UV absorptions (Tripp et al. 2000; Danforth et al. 2010, e.g.), X-ray absorptions (e.g. Fang et al. 2002; Kawahara et al. 2006; Buote et al. 2009; Fang et al. 2010) and X-ray emission lines by future surveys (e.g. Yoshikawa et al. 2003; Ohashi et al. 2006). It is therefore of prime importance to provide a tool which deals consistently with such possibly sparse discrete samples. Quite a few such options have been presented recently (Novikov et al. (2006); Hahn et al. (2007); Sousbie et al. (2008, 2009, 2008); Aragon-Calvo et al. (2008, 2010); Forero-Romero et al. (2009); Neyrinck (2008); Platen et al. (2007); Stoica et al. (2005, 2007, 2010), Sousbie (2010)), relying on different strategies on how to deal with these constraints (see Sousbie (2010)).

In paper I, the emphasis was on a formal presentation of the underlying mathematical theory and its extension to the discrete regime. As the corresponding algorithm is fairly intricate, a certain level of formal jargon was required to describe it unambiguously. Hence paper I focused on the language of mathematics. Conversely, let us first now rephrase here the corresponding framework in the more intuitive language of astrophysical data processing, as our aim is to appeal to both the community of computational geometry and that of astrophysics. What should be the expected characteristics of the ideal structure finder? Optimally, one would like to implement an algorithm which recovers the important and robust features of the underlying field even when little information is available, so that the procedure manages to reasonably identify the most striking features of the cosmic field. *Topology* (in fact discrete topology) therefore provides the natural context in which the optimal algorithm should be implemented. Indeed, topology *de facto* characterizes the “rubber” geometry of the underlying field, *i.e.* its most intrinsic and robust features. More specifically, as argued in Sousbie (2010), ideally such an optimal tool should produce an ensemble of critical sets (lines, surfaces and volumes) consistent with those defined within the context of Morse theory for sufficiently smooth fields (Milnor 1963; Jost 2008). Morse theory basically provides a rigorous framework in which to formally define such sets for “regular” density fields (roughly speaking regular means not de-

generate so that these sets are well defined). For instance, the critical lines defined by this theory connect peaks and maxima via special (extremal) flow lines of the gradient¹. These lines should trace quite well the filaments of the LSS. Similarly, the walls of the LSS should have a natural equivalent feature as the “critical” surfaces of Morse theory (the so called 2-manifolds).

Here our purpose is to proceed within the context of its discrete counterpart (Forman 2002). The corresponding discrete construction should be as consistent as possible with all the topological features of the underlying smooth field (it should globally preserve, at the level of the noise, the salient features of the field, such as the number of connected components, the number of tunnels or holes defined by its iso-contours, etc.; conversely² the significant discrete critical sets should have the correct “combinatorial” properties: *e.g.* critical lines should only connect at critical points, saddle points connect exactly two peaks together, etc...). The level of complexity of the corresponding network should also reflect the inhomogeneities of the underlying survey: *i.e.* it should adapt its level of description to the sampling, hence yielding a parameter-free multiscale description of the cosmic web. In fact, it should also provide a simple diagnostic in order to estimate the robustness of the various components of the network (*i.e.* the degree of reproduced details should be tunable to the purpose of the investigation). Finally it should clearly address the shortcomings of watershed-based methods described in paper I (namely the occurrence of spurious boundary lines induced by resampling in 3D or more).

Paper I presented precisely such a tool, while building up on an extension of Morse theory to discrete unstructured meshes. Two main challenges were addressed: (i) defining the counterpart of the (topologically consistent) critical sets on the mesh; (ii) defining a procedure to simplify the corresponding mesh at the level of the local shot noise.

The first step is achieved by considering simultaneously all the discrete components of the triangulated mesh (its vertices, edges, faces and tetrahedra), and reassigning a density to all these components in a manner which is heuristically consistent with the sampled density at the vertices; this relabeling procedure also ensures that the discrete flow (which follows from the corresponding discrete gradient) is sufficiently well-behaved to provide such topological consistency (specifically, it ensures the existence of discrete counterparts of regular critical points). Loosely speaking, amongst the discrete analogues of gradient flows, the algorithm identifies the critical subsets as special sets which cannot be paired to their neighbours through these discrete gradients. Note that the required level of compliancy to achieve this construction has the virtue of not only producing the discrete set of critical segments, but also the triangulated walls and voids at

¹ Indeed, Morse theory formalizes the process of partitioning space according to the gradient flow of the density into so called ascending and descending manifolds. In other words, it tags space according to where one would end up going “uphill” or “downhill”. In doing so it identifies special lines or surfaces, where something unusual happens.

² this well known duality between the topology of the level sets and the characteristics of the critical points clearly has a discrete analogue through the creation/destruction of discrete cycles, see paper I

no extra computational cost.

The second step follows from the concept of topological persistence (Edelsbrunner et al. 2000, 2002), which assigns a density ratio to pairs of critical points which are found to be connected together by such discrete integral paths; these pairs are identified by the destruction/creation of critical points as one describes the level sets. If this ratio is below a given threshold, the corresponding critical line/surface is found to be (topologically) insignificant, it is removed from the set and the remaining critical mesh is simplified so as to recover some topological consistency. In other words, if the shot noise induces the occurrence of the discrete counterparts of, *e.g.* spurious loops, disconnected blobs, or tunnels which are found to be insignificant according to the aforementioned criterium, they will be removed. The idea of topological persistence is central in producing a natural (topologically motivated) self-consistent criterion for inferring the significance level of the identified structures. In particular, it warrants that the removal of pairs of critical points consistently extracts the corresponding topological feature (loop, tunnel, component).

Importantly, let us emphasize that within this framework, the mathematical theories that we use are intrinsically discrete and readily apply to the measured raw data (modulo the consistent labelling of the elements of the Delaunay tessellation relative to the DTFE densities computed at the sampling points). This warrants that all the well-known and extensively studied mathematical properties of Morse theory are ensured by construction *at the mesh level*, and that the corresponding cosmological structures therefore correspond to well-defined mathematical objects with known mathematical properties.

In the language of computational geometry (see Appendix 5 for the relevant definitions), a simplicial complex (the tessellation) is computed from a discrete distribution (galaxy catalogue, N-body simulation, ...) using a Delaunay tessellation. A density ρ is assigned to each galaxy using DTFE (roughly speaking, the density at a vertex is proportional to the inverse volume of its dual Voronoi cell, see Schaap & van de Weygaert (2000)). A discrete Morse function (a re-labelling of all elements of the tessellation) is then defined by attributing a properly chosen value to each simplex in the complex (*i.e.* the segments, facets and tetrahedron of the tessellation). From this discrete function, we then compute the discrete gradient and deduce the corresponding Discrete Morse-Smale complex (DMC hereafter, Forman (2002)). The DMC (the set of critical points connected by arcs, quads, Crystals etc.) is used as the link between the topological and geometrical properties of the density field. Its critical points together with their ascending and descending manifolds (the “critical” sets) are identified to the peaks, filaments, walls and voids of the density field. The DMC is then filtered using persistence theory. For that purpose, we consider the Filtration (the discrete counter part of the density-sorted level-sets) of the tessellation according to the values of the discrete Morse function and use it to compute persistence pairs of critical points (pairs of critical points which are linked by their appearance and disappearance as one scans the Filtration). The DMC is simplified by canceling the pairs that are likely to be generated by noise. This is achieved by computing the probability distribution function of the persistence ratio (*i.e.* the ratio of the densities at the

connected pair) of all types of pairs in scale-invariant Gaussian random fields and canceling the pairs with a persistence ratio whose probability is lower than a certain level.

Paper I presented a couple of examples of such a construction in 2D. Let us now illustrate the virtue of the method in the context of the 3D cosmic Web. We start³ by showing that the geometry of the cosmic web is accurately reproduced, while illustrating the quality of the cosmological structures identification, both in an N-Body simulation (section 2) and directly on an unprocessed version of the SDSS DR7 galaxy catalogue (section 3). In particular we show how DisPerSE allows us to identify various configurations of the filamentary structure of galaxies, and identify a previously missed X-ray “optically faint” halo at the intersection of a set of SDSS filaments using the SUZAKU satellite. We then discuss in section 4 the problem of estimating the right value for the persistence level in cosmological simulations, and illustrate how the measured topological properties of the LSS distributions are affected by varying this threshold. In particular we show how this criterion compares with the simple friend-of-friend algorithm when attempting to identify halos in simulations. Section 5 wraps up and discusses prospects.

2 GEOMETRY OF LSS: SIMULATION

Although we have shown in paper I that the method introduced in that paper seems to be able to measure the topological properties of the cosmic web efficiently even in the presence of significant noise, we only illustrated in the 2D case that it could also recover correctly the geometry of the filamentary structure (see paper I). Demonstrating that a given algorithm is able to correctly identify the location of filaments is a difficult task, as it requires the previous knowledge of the location of those filaments. The only solution therefore seems to build an artificial distribution from a previously defined filamentary structure. This method was adopted in Aragon-Calvo et al. (2008), where the authors use a Voronoi Kinematic Model (van de Weygaert 2002). The principle of the Voronoi Kinematic Model is to identify the voids, walls, filaments and clusters to the cells, faces, edges and vertice of the Voronoi tessellation. In practice, randomly distributed particles are moved away from the nuclei of the Voronoi cells following a universal expansion rate, and their displacement being constrained to the faces, edges and finally vertice as they reach them. This results in a distribution of particles where each is said to be a void, wall, filament or cluster particle depending on whether they belong to a cell, face, edge or vertice of a Voronoi cell when the simulation is stopped.

We argue that using such a model to quantify the quality of the Morse-Smale complex identification is not as relevant as one would think, mainly because it is too idealized topologically speaking. In fact, it is a built-in property of the Voronoi Kinematic Models that all the cosmological structures overlap neatly: maxima (*i.e.* voronoi vertice) are located at the intersection of filaments (*i.e.*

³ Note that our goal here is not to present an exhaustive review of the geometrical properties of the cosmic web, which is clearly out of the scope of this paper.

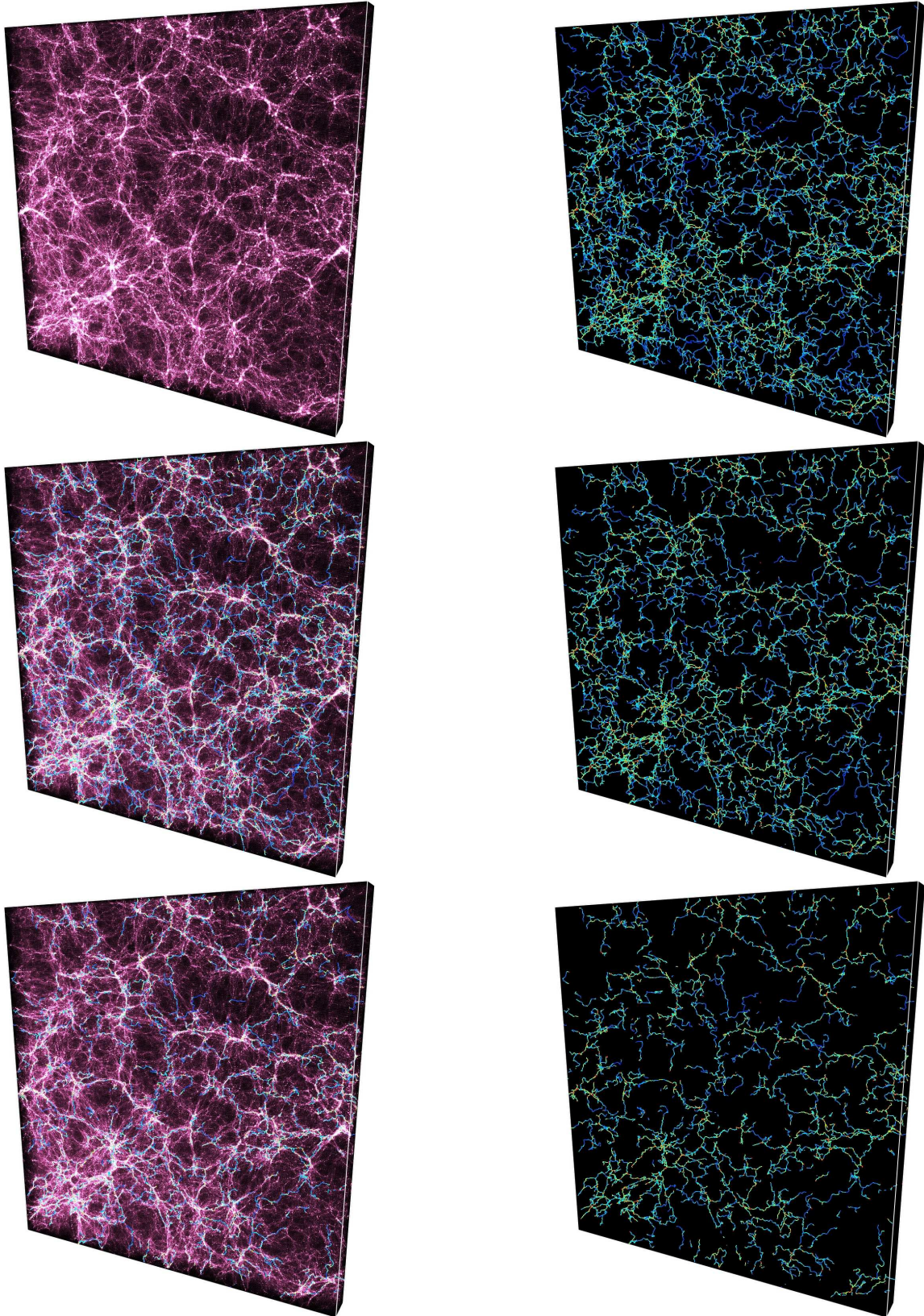
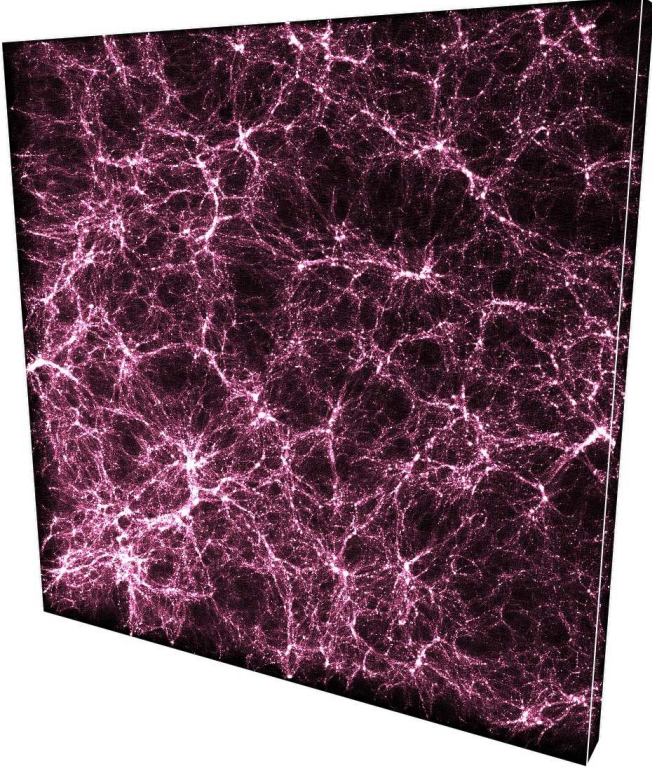
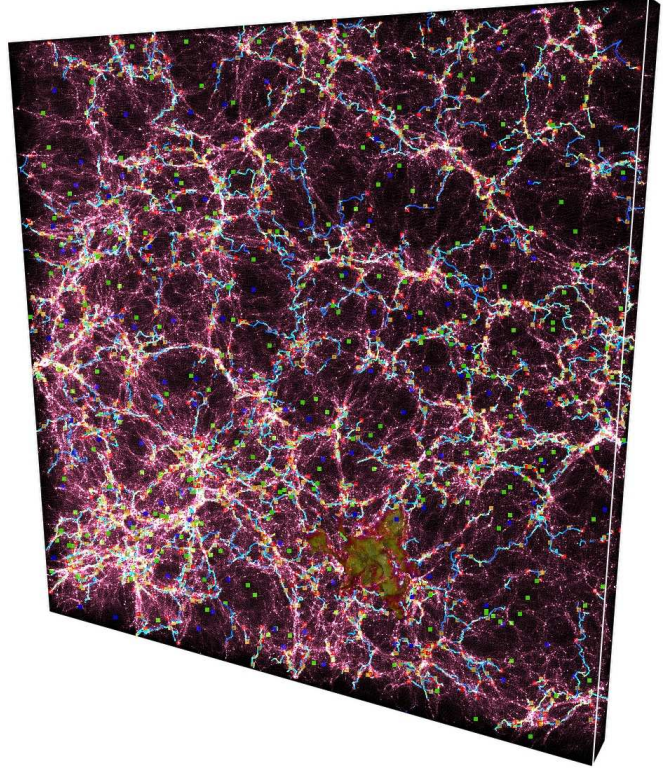


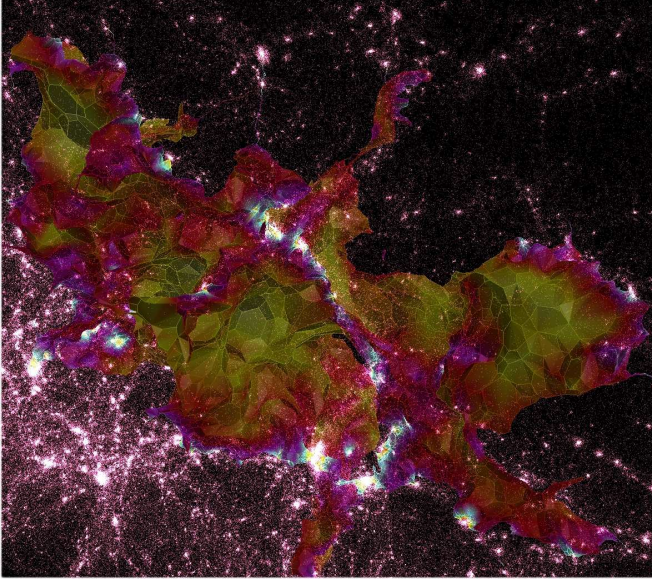
Figure 1. The filamentary distribution above a persistence level of $3\text{-}\sigma$, $4\text{-}\sigma$ and $5\text{-}\sigma$ (from top to bottom) in a $250 \times 250 \times 20 h^{-1}$ Mpc slice of a 512^3 particles and $250 h^{-3} \text{Mpc}^3$ large cosmological simulation. The computation was achieved on a 128^3 particles sub-sample, and the filaments are colored according to the logarithm of the density. The density field was represented using the 512^3 particles of the N-body simulation.



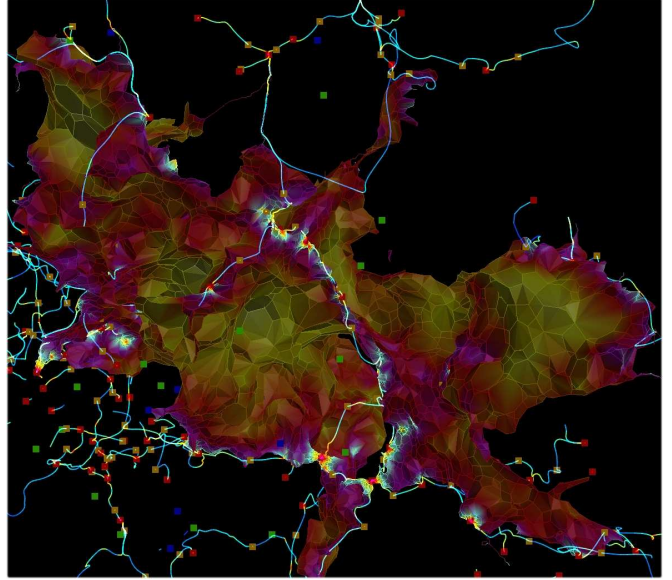
(a) Simulated dark matter distribution



(b) A void (bottom right) embedded in the filamentary structure



(c) Zoom on the void of panel 2(b).



(d) The relationship between the detected void, filaments and critical points

Figure 2. The arcs of the Discrete Morse-Smale complex (*i.e.* the filaments) and an ascending 3-manifold (*i.e.* a void) at a significance level of $5-\sigma$ in the same distribution as that of figure 1 (a $250 \times 250 \times 20 h^{-1}$ Mpc slice of a 512^3 particles and $250 h^{-3}$ Mpc 3 large cosmological simulation). The density distribution is represented using all available particles within the simulation (panel 2(a)) while the DMC was computed using 128^3 particles sub-sample. The 2 lower panels (2(c) and 2(d)) show a zoom on the upper panels at the location of the randomly selected void (see panel 2(b)). On these figures, the maxima, 1-saddle points, 2-saddle points and minima are represented as red, yellow, green and blue square respectively and the arcs as well as the manifold are shaded according to the log of the density. Note on panel 2(d) how the maxima, saddle-points and path of the filaments corresponds to the crests of the 2D density field measured on the surface of the void. This particularly emphasize the coherence of the detection of objects of different nature.

Voronoi segments) that always intersect with a suitable angle, those filaments are themselves by definition located at the intersection of at least three voids (*i.e.* voronoi cells), and each pair of neighboring void have exactly one Voronoi face in common, neatly defining the walls. As was shown in paper I, density functions extracted from actual data sets are in fact quite different, as they do not comply so easily to Morse conditions, in particular when measured from cosmological simulations or observational galaxy catalogues. In that case, and as clearly shown in paper I (see appendix 1), filaments may (and actually often do) merge before reaching a maximum, two apparently neighboring voids (down to the resolution limits) do not necessarily share a common face, and filaments are not necessarily at the intersection of at least three voids (once again, down to the resolution limit). The nature of the Voronoi Kinematic Model is therefore such that it avoids all the difficulty in identifying the Morse-Smale complex of realistic data sets. It might be possible to build more sophisticated Voronoi Models, that would for instance mimic the structure mergers that occur along the course of the evolution of large scale matter distribution in the Universe, but this is clear out of the scope of this paper. For the lack of a simple better way, we therefore use here what is probably to date the most efficient way to detect structures: the human eye and brain.

2.1 Visual inspection

The evolution of the geometry of the measured filaments with significance threshold is illustrated on figure 1. The DMC represented on this figure was computed at significance levels of 3, 4 and 5- σ (from top to bottom) within 128^3 particles sub sample of a 512^3 particles, $250 h^{-3} \text{ Mpc}^3$ ΛCDM dark matter only N-body simulation. Note that the dark matter distribution within the $20 h^{-1} \text{ Mpc}$ slice is represented in the top left corner to facilitate the visualization of its filamentary structure. Despite the projection effects that create visual artifacts (*i.e.* spurious filament looking structures resulting from the projection of dark matter clumps at different depths) and the fact that filaments may enter or leave the slice, therefore seemingly appearing and disappearing for no apparent reasons, it seems fair to recognize that the agreement between the observed and measured filaments is excellent. These good performances are mainly the result of our use of the scale adaptive Delaunay tessellation and the fact that our implementation does not require any pre-treatment of the density field, unlike usual grid based methods which enforce a maximal resolution and resort to some kind of density smoothing technique that affect the geometrical properties of the distribution. As a result, the resolution of the filaments is optimal with respect to the initial sampling whatever the selected significance level: the higher persistence and larger scale filamentary network is, by construction, a subset of its less persistent and lower scale counterpart. Because persistence based topological simplification is used, increasing the persistence threshold actually results in less significant filaments disappearing (when simplifying a 1-saddle point/2-saddle point persistence pair) or merging into each other (when simplifying a 1-saddle

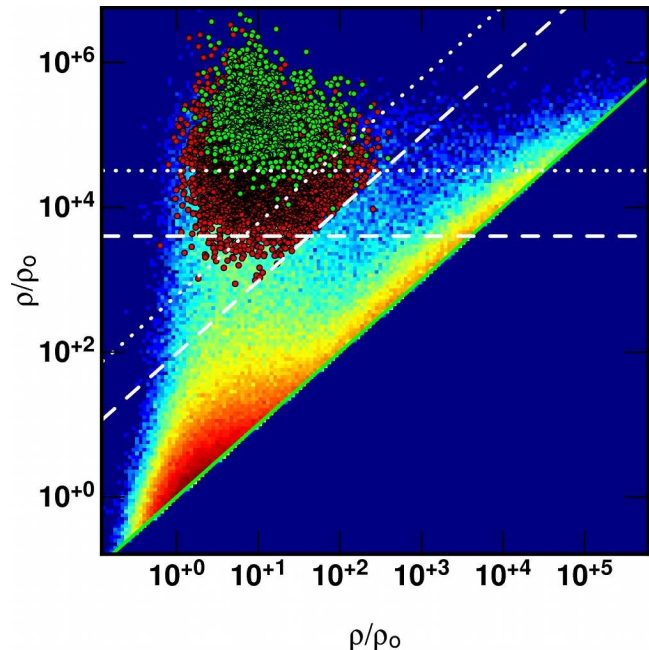
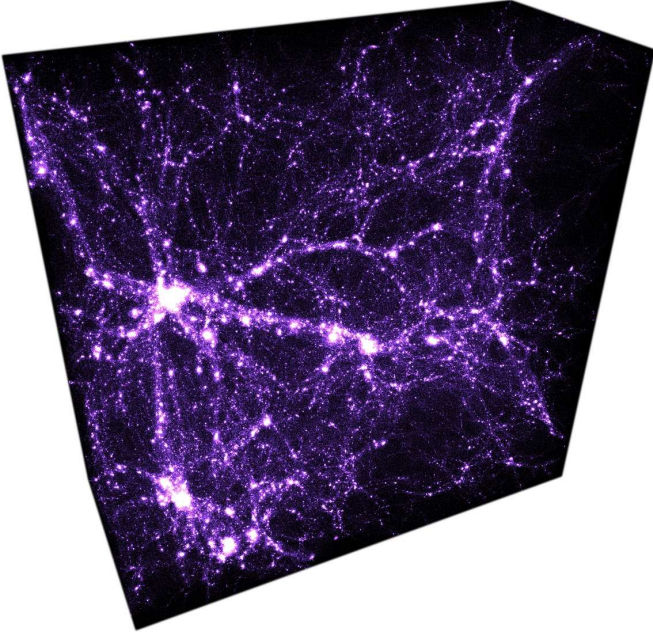


Figure 4. Distribution of the persistence pairs of the highest density particles within each dark matter halo of mass $M > 74 \times 10^{10} M_{\odot}$ (red) and $M > 590 \times 10^{10} M_{\odot}$ (green) in a 128^3 particle sub-sample of a $100 h^{-1} \text{ Mpc}$ large ΛCDM dark matter simulation. The persistence diagram of maxima/1-saddle-points pairs with persistence larger than $3-\sigma$ is shown in the background. The horizontal dashed and dotted lines correspond to overdensity levels of 4×10^3 and 3.2×10^4 respectively and the oblique lines to persistence levels of $\sim 4-\sigma$ and $\sim 5-\sigma$ respectively.

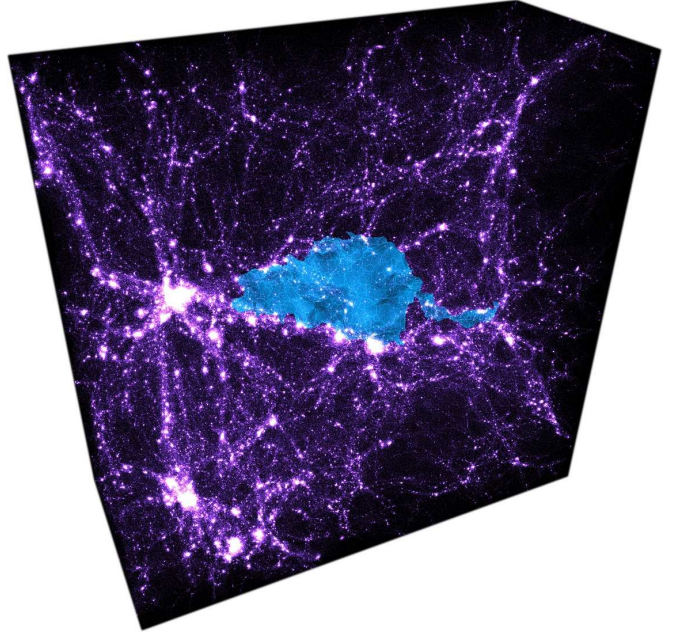
point/maximum persistence pair) to form larger scale more persistent ones, but conserving the exact same resolution in any case. This can easily be observed by comparing the filamentary networks on the right column of figure 1.

Another remarkable advantage of constructing cosmological structures identification on Morse theory is the extraordinary built-in coherence of the results, whatever the type of structure, as shown on figures 2 and 3. For instance, the intricate pattern of a randomly selected void (*i.e.* an ascending 3 manifold) embedded within the filamentary network (*i.e.* ascending 1 manifolds) of the cosmic web in the same simulation as previously is shown on figure 2. The location of the void within the slice is displayed on panel 2(b), each colored square standing for a critical point (see legend). On the zoomed frame (2(c) and 2(d)), the surface of the void has been shaded according to the logarithm of the density, showing how the DMC correctly traces the filamentary structure at the interface of the ascending 3-manifolds, as expected in Morse theory⁴. Similarly, the neat relationship between a detected void and a wall structure on its surface (*i.e.* an ascending 2 manifold) in a $100 h^{-1} \text{ Mpc}$ large N-body simulation is displayed on figure 3.

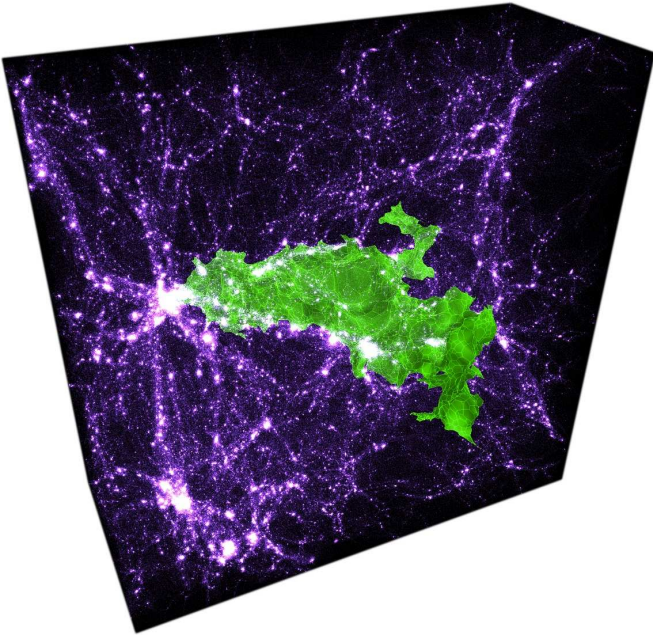
⁴ The slight shift in position between the surface of the void and the filament is due to the fact that we smoothed the filaments 4 times (see paper I)



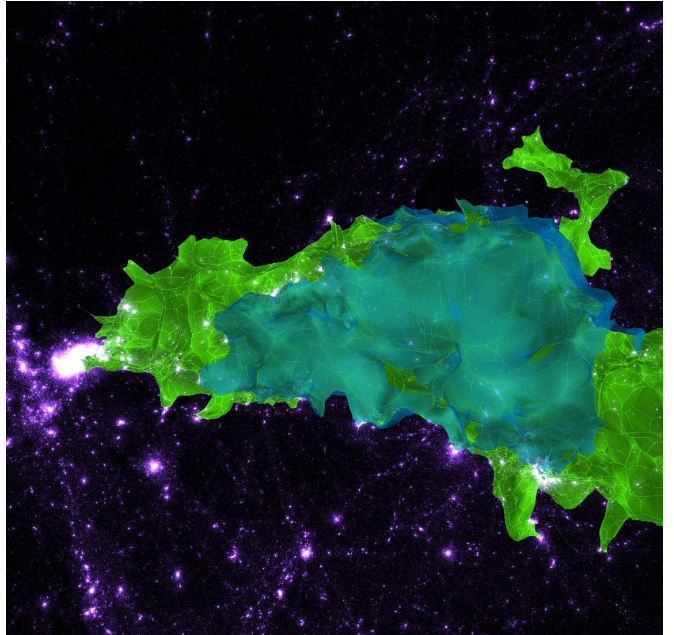
(a) Dark matter distribution in a $50 \times 50 \times 20 h^{-1}$ Mpc sub-box



(b) An ascending 2-manifold (*i.e.* a wall)

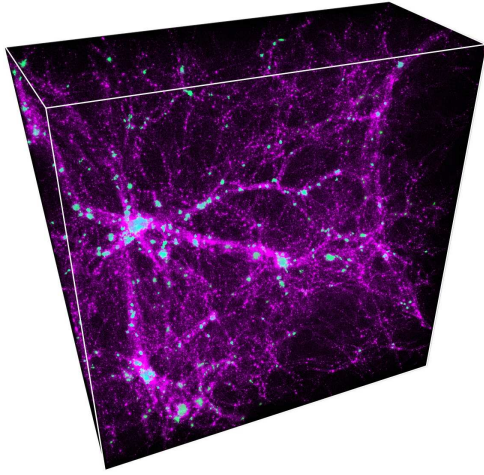


(c) An ascending 3-manifold (*i.e.* a void)

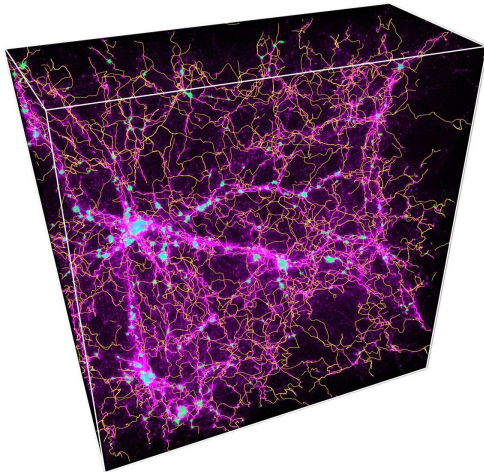


(d) Superposition of an ascending 3-manifold and an ascending 2-manifold on its surface.

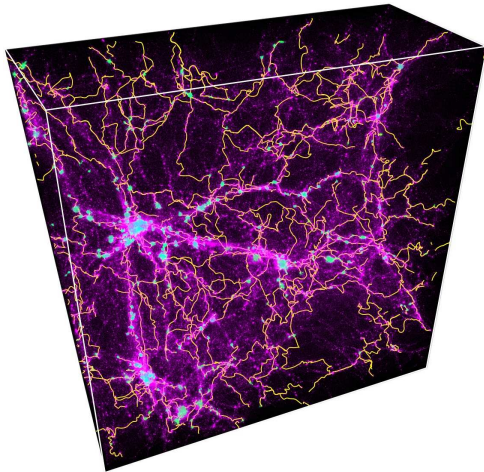
Figure 3. An ascending 2-manifold (*i.e.* blue 2D wall) and an ascending 3-manifold (*i.e.* green 3D void) identified in a 512^3 particles $100 h^{-1}$ Mpc Λ CDM dark matter simulation. The manifolds were computed from a 64^3 particles sub-sample.



(a) Dark matter distribution in a $50 \times 50 \times 20 h^{-1} \text{Mpc}$ sub-box and haloes with mass $M > 73.8 \cdot 10^{10} M_{\odot}$



(b) The filaments at $4\text{-}\sigma$ on a 128^3 sub-sample



(c) The main filaments of the dark matter haloes

Figure 5. Distribution of the main filaments of FOF haloes with mass $M > 73.8 \cdot 10^{10} M_{\odot}$ in a $20 h^{-1} \text{Mpc}$ thick slice of a 512^3 particles $100 h^{-3} \text{Mpc}^3$ ΛCDM dark matter simulation. The filaments were computed from a 128^3 particles sub-sample. Note that many filaments are linked to halos outside the slice, giving the false impression to end for no reason.

Let us finally address a straightforward question: to what extent does DisPerSE manage to grasp the main features of the cosmic web with relatively sparse samples? Figure 6 illustrates this query while comparing the filaments computed from two sub samples of varying resolution of from a $250 h^{-3} \text{Mpc}^3$ large cosmological simulation with 512^3 particles (namely 64^3 sub-sample and 128^3 sub-sample respectively). From this figure, it seems that indeed, the features which are identified in the sparser sample are real, since they are also found in the more densely sampled catalogue. There seems to be some encouraging level of convergence between the two sets of critical lines.

2.2 Persistent peak identification

From visual inspection, it therefore seems relatively clear that the technique developed in this paper is able to correctly decompose the cosmic web into simpler objects of astrophysical interest. However, the approach is based on one fundamental assumption, which is that the ascending and descending manifolds of Morse theory, each associated to a specific type of critical point, are representative of the voids, filaments, walls and haloes. While the astrophysical nature of a filament or a wall is not defined very precisely, but is rather understood intuitively, this is not the case of a dark matter halo for instance, which is supposed to be a gravitationally bound structure and the fact that the persistent maxima of the density field correctly identify the gravitationally bound structures is not established. We check this assumption by comparing the distribution of dark matter haloes identified by a simple friend of friend technique (FOF hereafter, see Summers et al. (1995) for instance) in a $100 h^{-3} \text{Mpc}^3$, 512^3 ΛCDM dark matter simulation to the persistence diagram in the same simulation, as illustrated on figure 4.

On this figure, the probability distribution function (PDF) of the persistence pairs⁵ of type 2 (*i.e.* the maxima/1-saddle points pairs) measured in a 128^3 particles sub-sample is displayed in the density/density plane, the horizontal axis corresponding to the density of the 1-saddle point, and the vertical one to that of the maximum. The green line therefore represents the 0-persistence limit, while the oblique white dashed and dotted lines delimits the $4\text{-}\sigma$ and $5\text{-}\sigma$ threshold respectively. In order to compare this distribution to that of the astrophysical dark matter haloes, each of them is also represented as a disk with coordinates that of the persistence pair of its most dense particle (the densest particle within a halo is necessarily a local maximum). Each halo was identified using a standard linking length parameter of one fifth of the average inter particular distance, and the red disks represent the haloes with mass $M > 73.8 \times 10^{10} M_{\odot}$ (*i.e.* with more than 1,280 particles in the initial simulation, or 20 in a 128^3 sub sample), while the green ones represent the haloes with mass $M > 590.4 \times 10^{10} M_{\odot}$ (*i.e.* with more than 10,240 particles in the initial simulation, or 1,280 in a

⁵ As in section 4, each pair is represented by a point with coordinates the density of each of its critical point, see that section for more explanations.

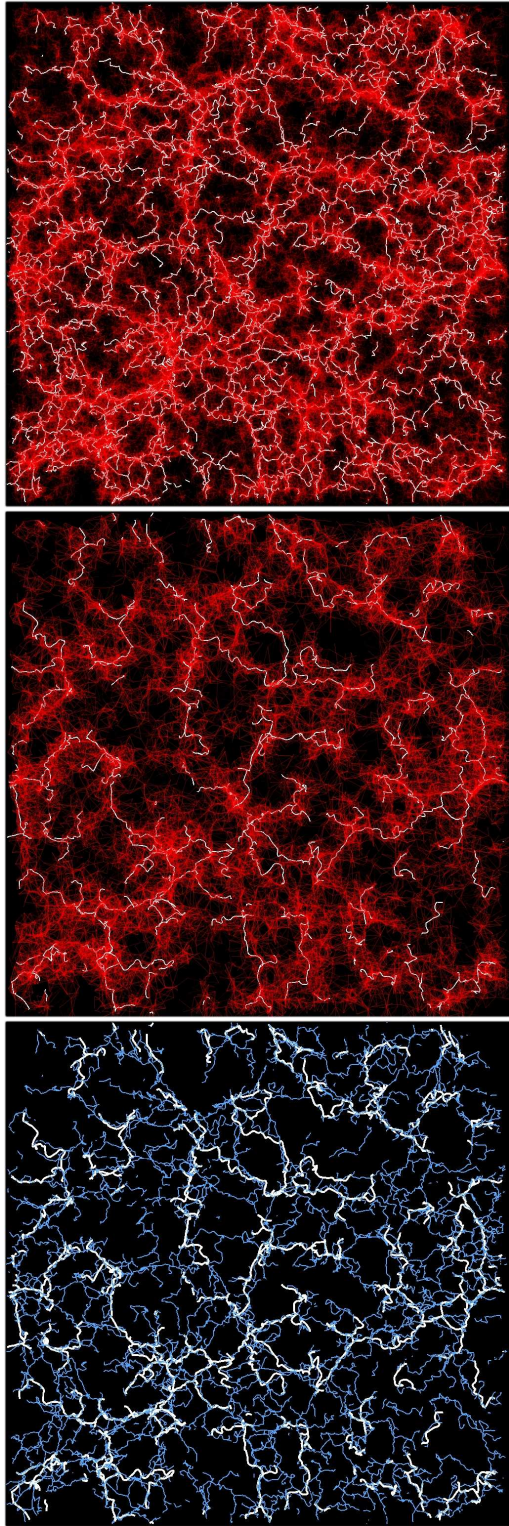


Figure 6. The filamentary distribution above a persistence level of $4-\sigma$ in a $250 \times 250 \times 20 h^{-1}$ Mpc slice of a 512^3 particles and $250 h^{-3}$ Mpc 3 large cosmological simulation. The red segments on the *top* and *central* figures correspond to the segments of the Delaunay tessellation of a 128^3 and 64^3 sub-sample, on which the corresponding filaments have been computed. On the *bottom* figure, the thick white filaments correspond to the 64^3 sub-sample while the blue thin filaments were computed on the 128^3 sub-sample. This figure clearly demonstrates that DisPerSE is able to grasp the main features of the cosmic web with relatively sparse sample.

128^3 sub sample). It is a very striking result how well the population of dark matter halos is localized in the persistence diagram. While lighter ones (red disks) mostly correspond to maxima with persistence ratio higher than $4-\sigma$ and overdensity $\rho/\rho_0 > 4 \times 10^3$, the heavier ones lie in the zone with persistence higher than $5-\sigma$ and overdensity $\rho/\rho_0 > 3.2 \times 10^4$.

These results mean that persistence selection associated to a global overdensity threshold is naturally (*i.e.* without any specific qualification) a very good halo finder, which is quite encouraging, and validates our initial assumption on the relationship between the persistent topological features and the astrophysical components of the cosmic web. This is illustrated on figure 5 where each dark matter halo with mass $M > 73.8 \times 10^{10} M_\odot$ (*i.e.* the red disks of figure 4) is colored in blue. Once again, it is clear on the central frame that all haloes along large filaments are correctly linked by the DMC. We also remark that the DMC and persistence pairs contain unexploited information of the topology as our algorithm explicitly identify the k -cycles as sequences of critical points associated to persistence pairs (see Sousbie (2010)). For instance, each persistence pair associated to a halo correspond to a 0-cycle that define a principal filament, as shown on the bottom frame, where only the filaments corresponding to persistence pairs whose maximum is a dark matter halo are represented. Moreover, using the information contained in the persistence pairs, one basically obtains a hierarchical structure finder that is able to also identify substructures not only within clusters, but also within filaments, walls and voids.

3 OUR UNIVERSE: THE SDSS CATALOGUE

Let us now illustrate a few prospective measurements of the filamentary structure of the actual galaxy distribution in the Universe. The ultimate goal of such measurements is to allow a complete and precise characterization of the properties of the filamentary structure of the galaxy distribution by measuring their topological properties, such as the Betti number and Euler characteristics, and modeling the geometrical characteristics of the voids, walls and filaments (*i.e.* their total length, number, the number of filaments per galaxy clusters, ...). Such a task is rather challenging, as it requires the construction of realistic mock observations from N-body simulations to assess the influence of observational biases and distortions; it also requires a lot of care in the handling of the observational data themselves (for instance by taking into account the complex survey geometry, among other difficulties). In this paper, we focus on convincing the reader that the method we introduced paper I is particularly suited to such a task by showing how easily and efficiently it can be applied to a real galaxy catalogue. We postpone the full investigation to a future paper.

3.1 The cosmic web in the SDSS

For that purpose, we use data from the 7th data release (DR7) of the SDSS (Abazajian et al. 2009), and in particular the large-scale structure subsample called *dr72bright0* sample of the New York University Value Added catalogue

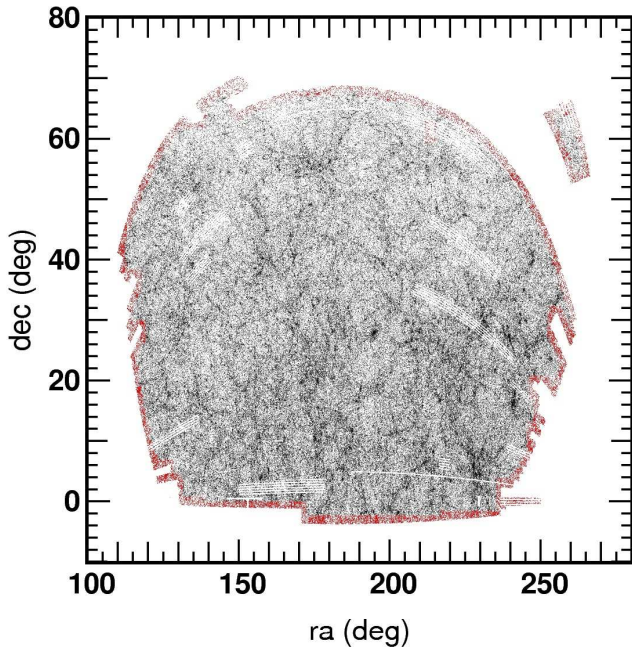


Figure 7. Angular distribution of the 515,458 galaxies corresponding to a sub-sample of the SDSS DR7 galaxy catalogue that we use in our tests (see main text for selection criterion). The 66,608 red galaxies are those detected as being on the boundary of the distribution using the method described in the main text. Note that some regions were not fully scanned and exhibit series of thin empty parallel stripes, but we simply ignore that fact when computing the boundaries.

(Blanton et al. 2005), which is made of a spectroscopic sample of galaxies with u,g,r,i,z- band (K-corrected) absolute magnitudes, r-band apparent magnitude m_r , redshifts, and information on the mask of the survey. In that sample, the spectroscopic galaxies are originally selected under the conditions that $10.0 \leq m_r \leq 17.6$ and $0.001 \leq z \leq 0.5$, but we further cut the sample for the purpose of our tests, restraining it to the galaxies with $z \leq 0.26$ and right ascension $100^\circ \leq RA \leq 280^\circ$, which removes the three thin stripes in the southern Galactic hemisphere. The resulting angular distribution, containing 515,458 galaxies among the 567,759 in the original sample is displayed on figure 7.

In order to compute the DMC of the observed galaxy distribution, we will use the mirror type boundary conditions as introduced in paper I. This type of boundary conditions normally apply to distributions enclosed within parallelepiped boxes, which is not the case here. In the simple case of a box-like geometry, the particles within a given distance of the faces are mirrored, and any particle outside the initial box or whose DTFE density may be affected by the content of the exterior of the box is tagged as a boundary particle. As the geometry of the SDSS catalogue is complex, we simply enclose it within a slightly larger box, fill the empty regions with a low density random distribution of particles, and then mirror the boundaries. The mirrored particles and the random ones are tagged and we then identify the boundaries of the galaxies distribution and tag as well those galaxies whose DTFE density may depend on the distribution outside the observational region.

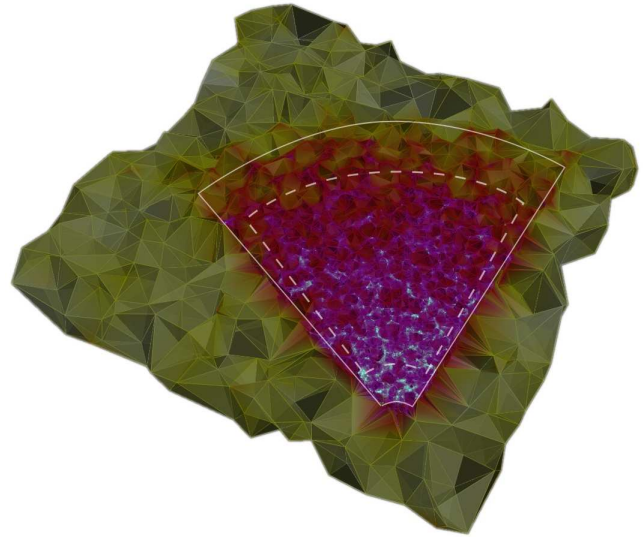
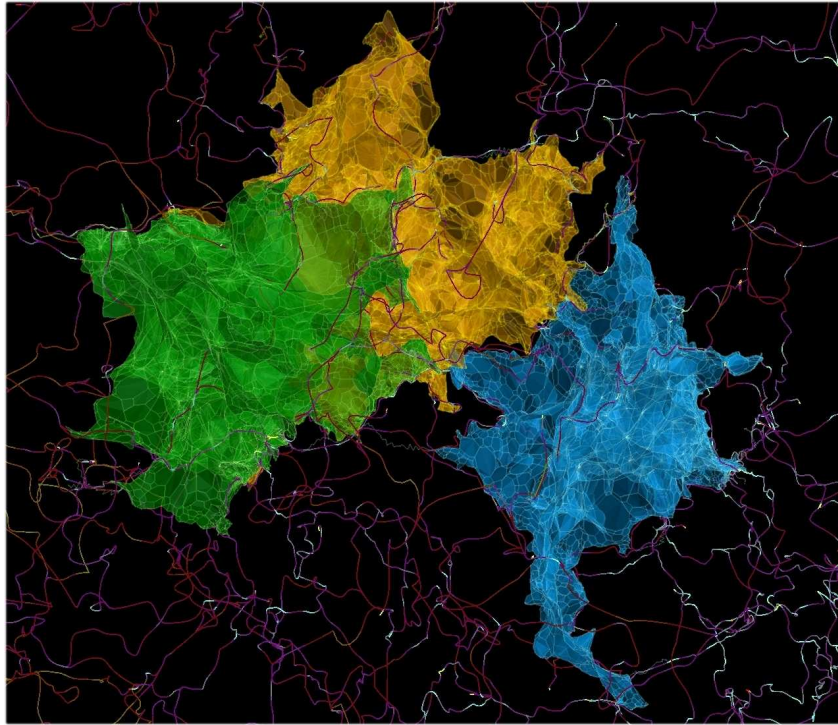
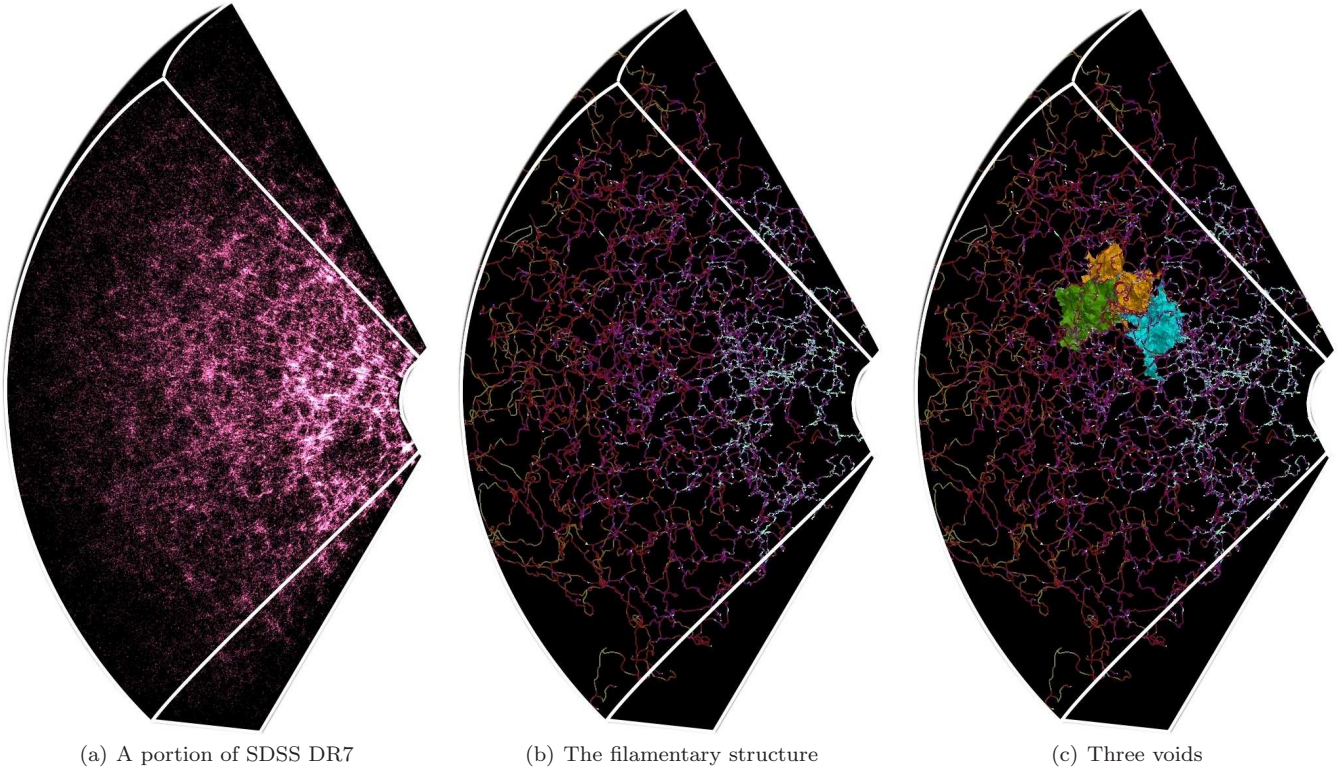


Figure 8. A slice within the delaunay tessellation of the distribution used to compute the DMC of the SDSS. The plain white contour delimits the SDSS distribution (inside) and the randomly added low density particles that fill the void regions of the bounding box (outside). Any galaxy outside the white dashed contours is considered as being on the boundary.

Although the catalogue does contain precise information about the mask of the survey, we prefer to use a simple though automatic method to identify the boundaries of the galaxy distribution. This method simply samples the angular galaxy distribution in the RA/DEC plane over a regular grid of $1 \times 1^\circ$ pixels, and identifies the galaxies on the boundary of the catalogue as those that belong to a pixel with at least one completely empty neighbor. Note that such a method presents the advantage of being generic, as it does not presume any previous knowledge of the mask, and could therefore be applied directly to other galaxy catalogues. The resulting boundary galaxies are represented in red on figure 7. We finally also tag those galaxies with redshift $z \leq 0.02$ and $z \geq 0.2$ as boundary and proceed with the computation of the DMC, as in the regular mirror type boundary condition case. A slice of the Delaunay tessellation of the final distribution is displayed on figure 8.

The resulting DMC covers the 440,950 galaxies in black on figure 7 and obeying the additional condition $0.02 \leq z \leq 0.2$ (or equivalently $85 \leq d \leq 860 h^{-1}$ Mpc) and it is displayed on figures 9, 10, and 11. Figure 9 illustrates the influence of the significance level on the measured filamentary network. On this figure, the filaments (*i.e.* the ascending 1-manifolds or arcs) within a $\sim 40 h^{-1}$ Mpc slice of the Delaunay tessellations are shown at significance levels of $3-\sigma$, $4-\sigma$ and $5-\sigma$ (from top to bottom); it is quite striking how well more or less significant filaments are accurately identified depending on the value of the persistence ratio threshold. Note how already at a level of $3-\sigma$ the influence of sampling noise has disappeared and increasing this threshold results in the selection of apparently denser, bigger and longer filaments. As the distant faint galaxies and the nearby bright ones cannot be



(d) A zoom on the voids and the filamentary structure

Figure 10. The detected filamentary structure at a significance level of $5-\sigma$ and three voids within a portion of SDSS DR7. Note that only the upper half of the distribution shown on figure 7 is displayed here for clarity reasons. The color of the filaments corresponds to the logarithm of the density field. The filaments of the SDSS extracted with DisPerSE is readily available online at the URL <http://www.iap.fr/users/sousbie/>.

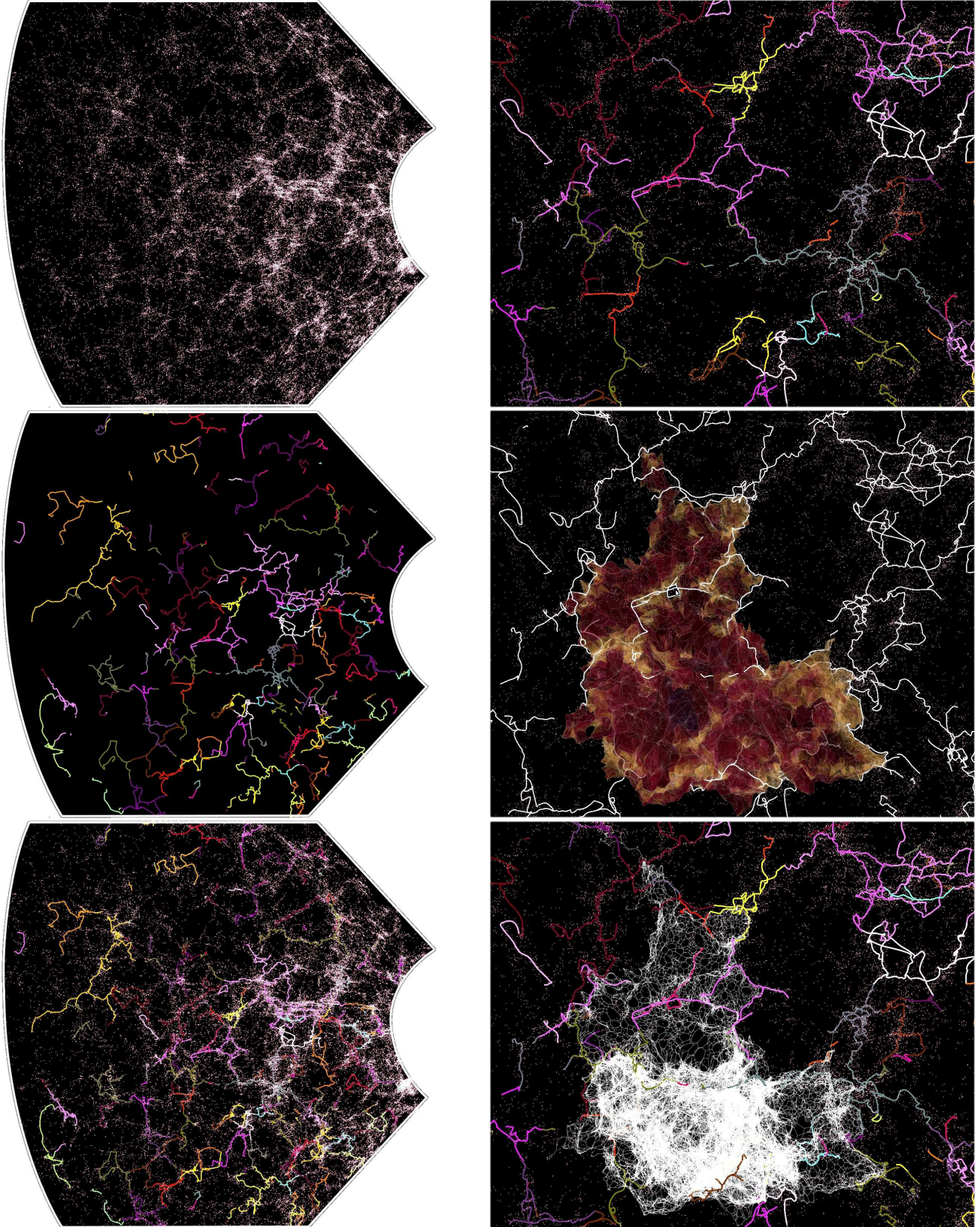


Figure 11. The filamentary structure (left) and a void (right) detected at a significance level of $5-\sigma$ in SDSS DR7. In order to emphasize the filamentary structure, only a $\sim 60 h^{-1}$ Mpc thick flat slice of the distribution is displayed on each frame. The void surface is shaded according to the log of the density field (central right frame), while the color of each arc of the DMC tracing the filamentary structure depends on the index of the maximum to which it is connected. Note that the foremost part of the voids on the central and bottom right picture protrudes from the slice, while the filaments are trimmed to its surface. Given its shape, this void is in fact a good example of why we should identify filaments via a DMC rather than using a Watershed technique, as it displays two strong “thin wings” which would lead to the incorrect detection of spurious sets of boundaries.

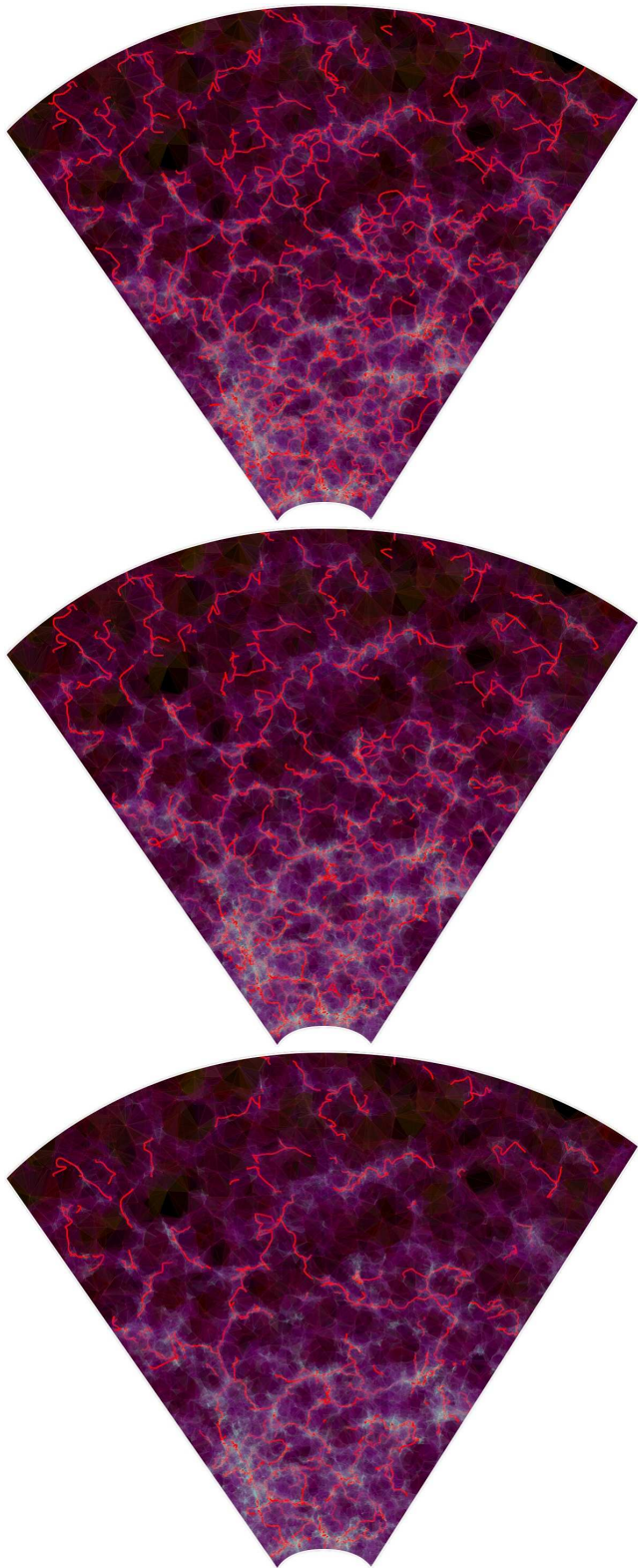


Figure 9. From top to bottom, the filamentary structure in a $\sim 40 h^{-1}$ Mpc thick slice of the SDSS Dr7 galaxy catalogue at a significance level of 3, 4 and $5-\sigma$ respectively. The distribution is represented by the non bounding subset (see main text) of the Delaunay tessellation used to compute the DMC, shaded according to the logarithm of the density. The depth of a filament can be judged by how dimmed its shade is. Note that filaments that seem to stop for no apparent reason actually enter or leave the slice.

observed easily, the selection function strongly depends on the distance, and so does the sampling. It reflects in the shade of the Delaunay tessellation, which depends on the logarithm of the density. From a theoretical point of view, the fact that the absolute value of the density is multiplied by the selection function should not affect the detection of the filaments as long as the value of the selection function does not vary much over the typical scale of a filament (or in other word, as long as the topology of the distribution remains unchanged). The measured persistence ratio of persistence pairs may be slightly affected though, when the two critical points in the pair are located at different distances, but this does not seem to have much importance in the present case. A more significant effect results from the scale adaptive nature of DTFE. Because the quality of the sampling decreases with distance, comparatively larger scale filaments are identified as the distance increases and to be able to identify comparable filaments independently of the distance from the observer, one would therefore probably have to resort on volume limited samples.

The filamentary structure at $5-\sigma$ significance level is also shown over larger scales on figure 10 and within a $60 h^{-1}$ Mpc slice where each galaxy is represented by a point on figure 11. Three voids (*i.e.* ascending 3-manifolds) have been randomly selected within the distribution of figure 10 and are displayed on the bottom frame 10(d), showing the intricate relationship between the voids and the filamentary structure that crawls at their surface. As previously observed in simulations, it can be seen on the central right frame of figure 11 that those 3D filaments also trace the 2D filamentary structure at the surface of the voids as expected from Morse theory. Note that it is only because they have been smoothed over four segments to look more appealing and to avoid rendering problems that the filaments do not lie precisely on the surface of the voids. It is in fact a build-in feature of the DMC and in particular of our implementation that all the different types of identified cosmological structures do form a coherent picture, whatever the properties of the initial discrete sample. This allows for interesting features, such as making possible the count of the number of filaments that belong to a common maxima by intersecting the ascending 1-manifolds with the descending 3-manifolds. This is shown on figure 11 where the color of the filamentary structure corresponds to the index of the maximum it belongs to and individual filaments could be identified the same way, as the two arcs of the DMC originating from a given saddle point.

3.2 An “optically faint” cluster at a filamentary junction

Because some dark matter haloes are sparsely populated and also as a result of selection effects, classical methods such as FOF are unable to detect them from the observed galaxy distribution. Such “optically faint” groups and clusters may nevertheless present a strong astrophysical interest: as they are by nature different from the “regular” haloes, one could for instance expect that they have different formation history that needs to be understood. As they are faint though, their properties are poorly assessed, but massive dark matter haloes such as galaxy clusters or

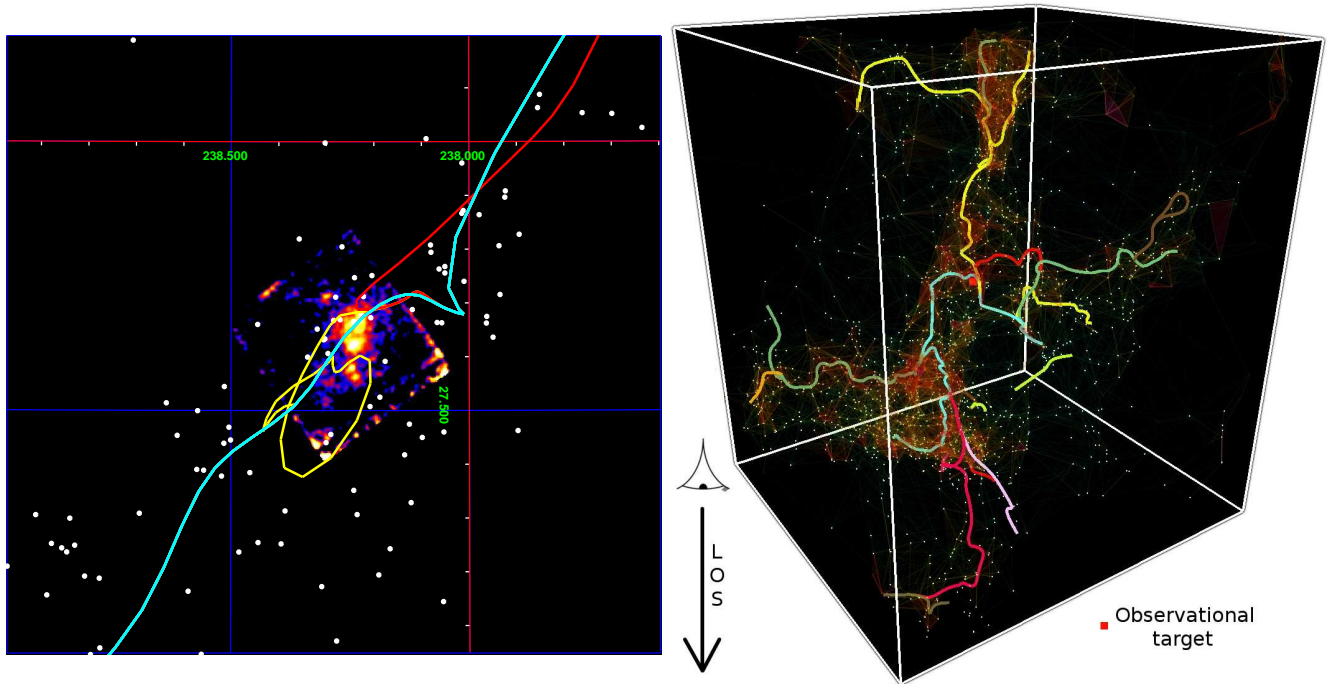


Figure 12. Left: An X-ray halo observed around an elliptical galaxy in the center of a group at redshift $z = 0.083$ and located at the confluence of several filaments. The color map indicates the X-ray combined image of CCD chips (XIS 0,1, and 3), while the white dots stand for the SDSS spectroscopic-identified galaxies within $0.080 < z < 0.086$. The filamentary structure in the surrounding region is shown by the colored solid curves, extracted from the filaments catalogue shown on figure 10(b). Note that the colors (cyan, red and yellow) correspond to that of the filaments represented on the 3D view on the right frame. Right: a 3D view of the configuration of the filaments around the observed region. The vertical axis corresponds to the line of sight (the observer being upward), and the box roughly encompasses the galaxies in the SDSS catalogue with coordinates $233^\circ < \text{DEC} < 243^\circ$, $22^\circ < \text{RA} < 32^\circ$ and $0.075 < z < 0.092$. The delaunay tessellation of the galaxies, shaded according to the local density, is displayed to help visualizing the filamentary structure. The observational target is identified by a red square and is located at the intersection of the red cyan and yellow filaments, the last two being aligned with the line of sight to a very good approximation. A movie is available for download at <http://www.iap.fr/users/sousbie/>.

galaxy groups are believed to form at intersections of two or several filaments, which can be identified in the SDSS using DisPerSE. We demonstrate that this is possible by enlightening the relationship between an X-ray halo and its surrounding filamentary network as identified in the SDSS catalogue (see figure 10(b)).

Because of the particular configuration of the filaments in the region, we submitted an observation proposal to the X-ray satellite SUZAKU (Mitsuda et al. 2007), which was accepted. We present here the results of this observation, but reserve its analysis to a future article (Kawahara et al 2010, *in prep.*). The observational target was selected for being located at the confluent of galaxy filaments, and because one of those filaments is both straight and aligned with the line of sight as shown on figure 12 (see the yellow filament on the right frame). While no X-ray signal could be found within the ROSAT All Sky Survey (RASS), X-ray signals emitted by diffuse thermal gas were clearly observed by the high sensitivity detectors of SUZAKU, unveiling the presence of a dark matter halo as shown by the X-ray image reproduced in the central part of the left panel of figure 12. It is remarkable that there are no corresponding candidates in the 78,800 groups catalogue identified by (Tago et al. 2010) using a modified friend-of-friend (FOF) algorithm. In fact, because the optically observable member galaxies are not strongly clustered and their number is limited ($N \sim 10$),

regular methods have high chances to miss them. It is also very difficult to locate and identify particular filamentary configurations by eye directly from the galaxy distribution using projections or even a real time 3D visualization. Using DisPerSE, we showed that it is possible to easily identify such targets, which demonstrate the complimentary of our approach with respect to one based on a traditional halo finders.

4 SIGNIFICANCE OF TOPOLOGY OF LSS

As noted in paper I, it is not an option to use the raw Discrete Morse-Smale complex as a tool to assess the properties of the cosmic web. Hence we showed there how to simulate a topological simplification of the DTFE density field so that the critical simplexes that were most probably accidentally generated by Poisson noise could practically be removed from the DMC. This simplification is based on the persistence ratio of critical points pairs (*i.e.* persistence pairs), and one must therefore decide a significance level $s = n - \sigma$ such that all persistence pairs with lower significance (*i.e.* or equivalently a higher probability to be generated by Poisson noise) can be removed. We showed in paper I that, at least in the 2D case, such a method allows for what seems to be a very efficient and natural simplification of the DMC. We did not discuss however

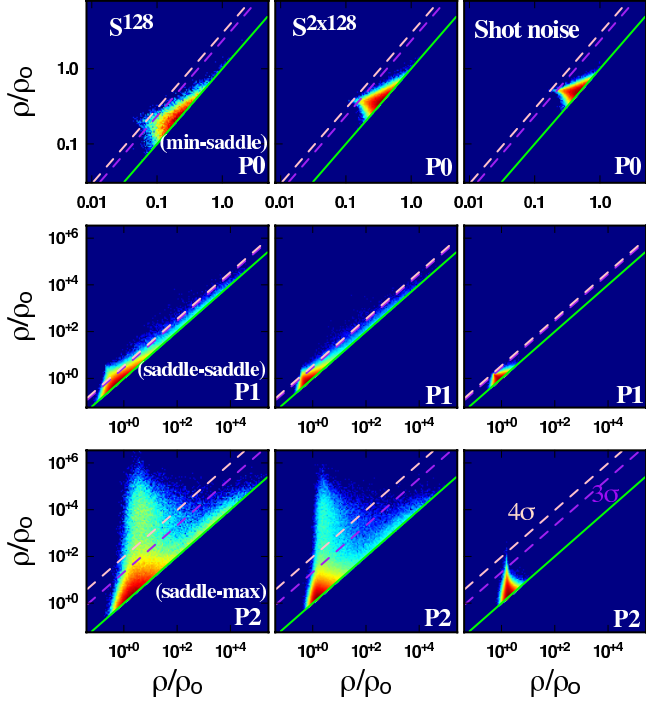


Figure 13. persistence diagrams (*i.e.* the probability distribution function (PDF) of persistence pairs) in a cosmological simulation and for Gaussian random noise. Each pair $P_i = [p_i, q_{i+1}]$ of critical points of order i and $i + 1$ is considered as a point with coordinates $[\rho(p_i), \rho(q_{i+1})]/\rho_0$. The PDF were computed from a $250 h^{-1}$ Mpc large Λ CDM dark matter simulation down sampled to 128^3 particles, S^{128} (left column), the same distribution with 128^3 additional randomly located particles, $S_N^{2 \times 128}$ (central column), and a random distribution of particles within the same volume, S_R^{128} (right column). From top to bottom, each line correspond to a different type of pair: P_0 (minima/2-saddle points), P_1 (2-saddle points/1-saddle points) and P_2 (1-saddle points/maxima) respectively. The green, purple dashed and pink dashed lines correspond to $0-\sigma$, $3-\sigma$ and $4-\sigma$ persistence levels respectively.

how to decide the value of this particular threshold. This is particularly important though, and especially in the context of the cosmic web, as our ultimate goal is to assess physical properties of astrophysical objects identified as features of the DMC (*i.e.* the haloes, filaments, walls and voids of matter distribution on cosmological scales in the Universe). Imagine for instance one is interested in statistically measuring the average number of filaments that branch on dark matter halos. If the threshold is too low, the measure will be equivalent to that in a Gaussian random field because of Poisson noise (see lower left frame of figure 13 of paper I), and if it is too high, then the risk is to systematically ignore weaker filaments (see central right panel of figure 13 of paper I).

4.1 Persistence diagrams

Figure 13 shows the probability distribution function of persistence diagrams (see Edelsbrunner et al. (2000), Cohen-Steiner et al. (2007)) computed from the Delaunay tessellation of a $250 h^{-1}$ Mpc large, 512^3 particles Λ CDM

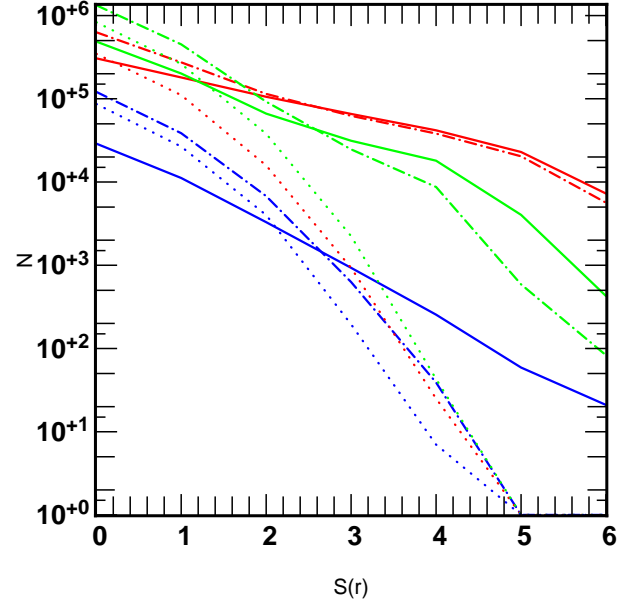


Figure 14. Number of persistence pairs of type k as a function of the significance threshold $S_k(r)$ (in units of σ) in a $250 h^{-1}$ Mpc large Λ CDM dark matter simulation down sampled to 128^3 particles, S^{128} (filled curves), the same distribution with 128^3 additional randomly located particles, $S_N^{2 \times 128}$ (dash-dotted curve) and a random distribution of particles within the same volume, S_R^{128} (dotted curves). The blue, green and red color correspond to persistence pairs of type 0, 1 and 2 respectively (see figure 13 for the corresponding persistence diagrams).

dark matter simulation subsampled to 128^3 particles (left column, S^{128} hereafter), the same distribution with an identical number of particle added at random locations (central column, $S_N^{2 \times 128}$ hereafter), and a completely random distribution of particles within the same volume (right column, S_R^{128} hereafter). Simply speaking, plotting a persistence diagram of a density distribution ρ basically consists in representing each persistence pair $P_i = [p_i, q_{i+1}]$, where p_i and q_{i+1} are critical points of order i and $i + 1$ respectively, as a point of coordinates $[\rho_\downarrow, \rho_\uparrow] = [\rho(p_i), \rho(q_{i+1})]/\rho_0$ where ρ_0 designates the average density in the distribution⁶. Recall that persistence pairs are pairs of critical simplexes that correspond to the act of creation and destruction of a topological feature in the Filtration of the Delaunay tessellation. On figure 13, the pairs of type P_0 , P_1 and P_2 are represented on the top, central and bottom rows respectively. On those diagrams, the pairs with null persistence lie on the green line of equation $\rho_\uparrow = \rho_\downarrow$ and the farther away from this line a point is, the higher the persistence of its corresponding persistence pair. The purple and pink dashed line stand for $3-\sigma$ and $4-\sigma$ persistence respectively. As expected, most persistence pairs in the random distribution S_R^{128} have a persistence ratio below $3-\sigma$ (right column). Fortunately, the PDF of the persistence pairs in S^{128} is sufficiently different from that in S_R^{128} so that a reasonable fraction of

⁶ In the following, the term density will generally refer to the normalized density ρ/ρ_0 so that different distributions can be fairly compared

them lie above the $3-\sigma$ and even $4-\sigma$ threshold (see left column). By canceling all those pairs that lie below the $3-\sigma$ or $4-\sigma$ line, it should therefore seem reasonable to assume that only those topological properties that were imprinted by the physical processes at work in the simulation would be conserved. A good measure of the actual influence of Poisson noise on the distribution of the persistence pairs in the underlying distribution can be gained from the examination of the central column. The distribution $S_N^{2 \times 128}$ was created by adding a large number of randomly located particles to S^{128} , resulting also in the creation of a very large number of spurious critical points. One can see on the central column that as a result, the persistence diagram tends to concentrate at lower persistence ratio (*i.e.* closer from the green line). This means that as expected, those spurious critical points mainly create low persistence ratio pairs which can therefore be removed.

This observation is supported by figure 14, where the actual number of persistence pairs in the three distributions are displayed as a function of the cutting threshold. Whereas the number of critical pairs of all sorts and with significance higher than $0-\sigma$ is higher in $S_N^{2 \times 128}$ (dash-dotted curves) than in S^{128} (plain curves), this number decreases comparatively faster with the increase of the persistence selection threshold. For low persistence thresholds (*i.e.* up to $\sim 2-\sigma$), the number of persistence pairs in $S_N^{2 \times 128}$ actually decreases as fast as that in the random distribution S_R^{128} (dotted curves). In the case of pairs of type P_1 and P_2 (2-saddle points/1-saddle points pairs, green curves, and 1-saddle point/maxima pairs, red curves, respectively), this tendency actually changes between $2-\sigma$ and $3-\sigma$ and the cancellation rates in $S_N^{2 \times 128}$ and S^{128} become relatively similar above $3-\sigma$. This strongly suggests that most of the spurious persistence pairs in $S_N^{2 \times 128}$ do in fact have a persistence ratio lower than $3-\sigma$ and that above that threshold, the remaining persistence pairs have a distribution similar to that in the original N-body simulation S^{128} . The persistence pairs of type 0 in $S_N^{2 \times 128}$ (minima/2-saddle point pairs, blue filled curves) exhibit a slightly different behavior though, as their number seems to vary more or less accordingly with the persistence threshold in $S_N^{2 \times 128}$ and S_R^{128} (blue dotted curve). This number nevertheless always remain higher in $S_N^{2 \times 128}$ and there are proportionally more high persistence pairs in $S_N^{2 \times 128}$ than in S_R^{128} . This suggests that the number of minima resulting from the physical processes at stake in voids formation is relatively low compared to that due to Poisson noise, the reason for this being that the cosmological voids' minima have an intrinsically lower density because of the nature of voids. While Poisson noise creates spurious minima over a wide range of densities, the voids' minima only span the lower densities and therefore stretch over comparatively larger scales due to DTFE properties (resolution being inversely proportional to the density). The addition of random particles in $S_N^{2 \times 128}$ particularly affects the wider regions around minima, therefore increasing their density and lowering the persistence ratio of the corresponding persistence pairs, hence the lack of high significance pairs of type 0 at $S(r) > 5-\sigma$ (see blue curves) in $S_N^{2 \times 128}$ compared to S^{128} . Note however that this does not mean that the physically created persistence pairs are destroyed

by Poisson noise in $S_N^{2 \times 128}$, but only that they are shifted to lower persistence, and that the persistence threshold should not be chosen too high if ones wants to retrieve the full DMC (which is not the case if one is only interested in the filaments).

Two complementary measures of the evolution of the topological properties in S^{128} and $S_N^{2 \times 128}$ with the persistence threshold are presented on figure 15: the PDF of the critical points on figure 15(a) and the betti numbers and Euler characteristics on figure 15(b).

4.2 Critical points

Let us consider figure 15(a) first. On that figure, the PDF of the density at vertice (*i.e.* the particles in the studied distribution) is shown by the dark black bold curve, and it is striking how the PDF of the critical points tend to follow it, especially at low persistence (outer curves): the more the k -simplexes at a given density level, the higher the number of detected critical points of order k . This is an expected result when Poisson noise dominates as it affects indifferently any scales, but it is not desirable though as the filamentary structure of the cosmic web is an intrinsic property which should not depend on the properties of a particular sampling technique. One would in fact rather expect the PDF of the critical points to follow the PDF of the volume weighted density, or equivalently as we use DTFE, of the number of vertice at a given density in the tessellation⁷. The black bold dashed curve traces the *volume weighted* PDF of the density at vertice. It is clear on figure 15(a) that in the case of the minima, 1-saddle points and 2-saddle points PDFs, the bias toward higher better sampled densities due to DTFE is progressively wiped out with increasing persistence ratio threshold, and almost disappears above a significance level threshold of $\sim 3-\sigma$ (see blue, green and purple curves). The PDF of the maxima though (red curves) exhibits an opposite tendency, as their PDF concentrates at higher and higher densities with increasing persistence ratio thresholds. This actually reflects the nature of the distribution of the dark matter over large scales in the universe. In fact, most maxima are expected to be found within gravitationally bound structures undergoing non-linear regime (*i.e.* dark matter haloes), which therefore exhibit densities several order of magnitudes higher than the average density and with very steep gradients (note that this fact also prevents them from being affected by Poisson noise too much). Those regions, although numerous, represent only a very small fraction of the total volume, as reflected by the discrepancy between the PDF of the maxima at high persistence ratio and the volume weighted PDF of the density. To confirm these hypothesis, we traced on figures 15 and 15(b) the blue and red vertical dotted lines which mark the characteristic average under-density of a void in a Einstein-de Sitter model,

⁷ in the case of DTFE, the density of a sample particle is defined as the inverse volume of its dual Voronoi cell, and the volume it occupies is also the volume of this cell, which implies that the PDF of the volume weighted density and that of the number of sample particles are identical.

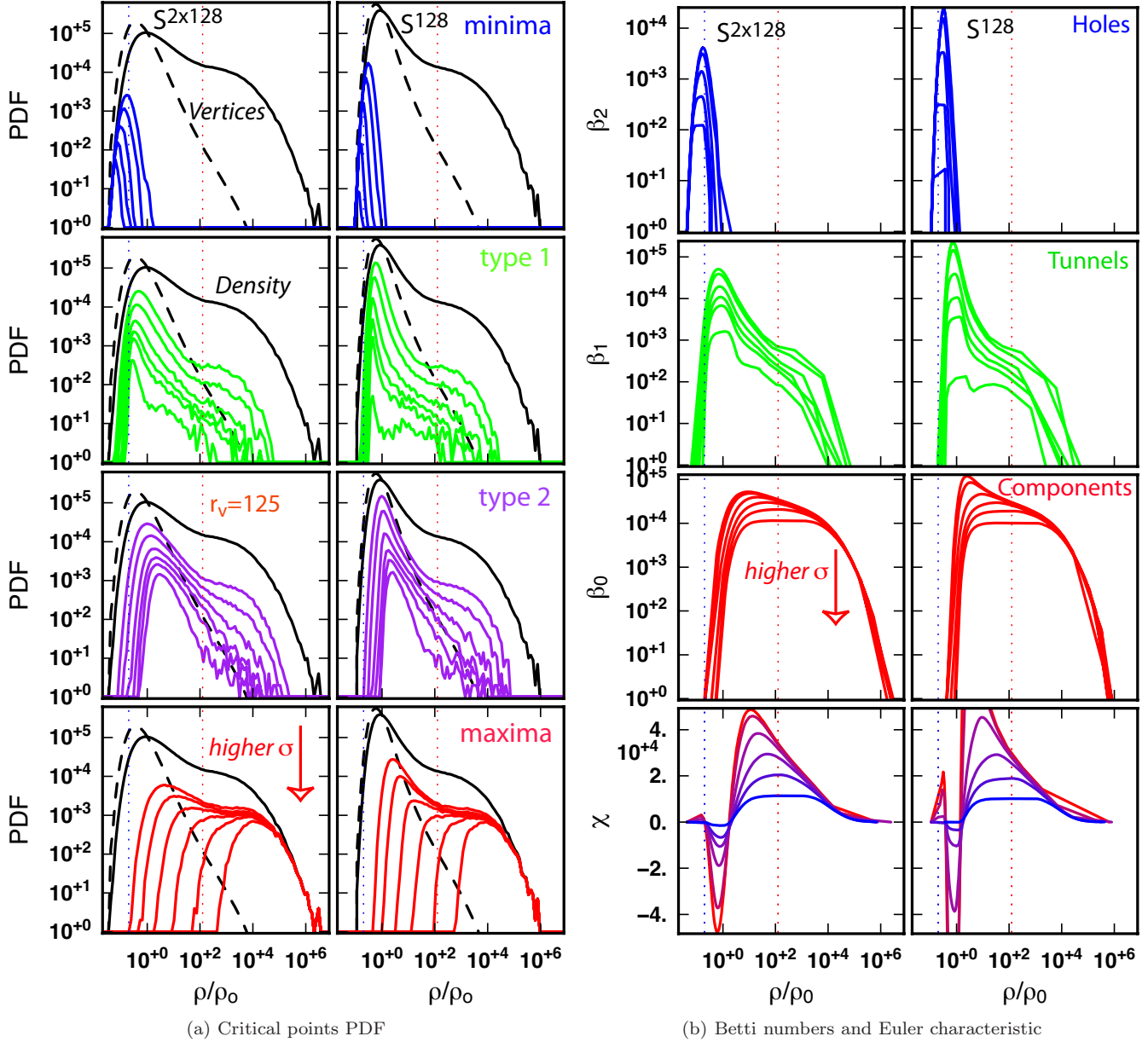


Figure 15. Evolution of the topological properties in a 512^3 particles $250 h^{-1}$ Mpc dark matter simulation down-sampled to 128^3 particles, S^{128} , for increasing persistence levels (left columns on each figure), and in the same distribution with 128^3 additional randomly located particles, $S^{2 \times 128}$ (right columns on each figure). On each frame, the persistence selection level ranges from $0-\sigma$ for the outer colored curve up to $6-\sigma$ for the inner curve. *Left:* The probability distribution function (PDF) of critical points of type 0 (top) up to 3 (bottom) as a function of their overdensity ρ/ρ_0 . The black curve is the PDF of the vertex in the tessellation while the dashed curve stands for the (volume weighted) PDF of the overdensity ρ/ρ_0 . The blue and red vertical dotted lines emphasize the critical level $r_v = \rho_v/\rho_0 = 0.2$ (resp. $r_p = \rho_p/\rho_0 = 125$) below (resp. above) which a void (resp. a peak) may be considered physically significant. *Right:* from top to bottom, the betti numbers, β_2 , β_1 , β_0 , and Euler characteristic χ of the excursion set with over density greater than ρ/ρ_0 .

$\rho/\rho_0 \leq 0.2$ (see Blumenthal et al. (1992), Sheth & van de Weygaert (2004) or Neyrinck (2008)) and the typical critical overdensity above which gravitationally bound structure are identified using friend of friend algorithm, $\rho/\rho_0 \geq 125$ (Summers et al. 1995) respectively. While this is not clear at low persistence thresholds because of Poisson noise, all maxima (resp. minima) belonging to persistence pairs with persistence ratio greater than $\sim 3-\sigma$ have densities above (resp. below) those critical thresholds while the two types

of saddle points lie within those limits. This means that the detected persistent maxima and minima correspond to physically meaningful objects, which strongly supports the pertinence of using persistence based cancellation of a Morse-Smale complex to identify the characteristics components of the cosmic web such as cosmic voids and filaments.

4.3 Discrete topological invariants

The Betti numbers and Euler characteristics represented on figure 15(b) are slightly more involved topological analysis tools than the PDF of critical points (see paper I for a more formal definition of the Betti numbers and a simple example of their computation). The k^{th} Betti number β_k counts the number of k -cycles in excursion sets as a function of the density threshold of the excursion. Within the context of the 3D cosmological matter distribution, there are 3 Betti numbers, that count the number of holes or 2-cycles (β_2), the number of tunnels or 1-cycles (β_1) and the number of distinct components or 0-cycles (β_0) enclosed in the set of points with density threshold larger than the aforementioned density threshold. As this threshold decreases, new k -cycles may be created or destroyed, therefore increasing or decreasing the value of the corresponding Betti numbers. The value of the Betti numbers as a function of the density threshold reflects the global topology of the field (*i.e.* the way it connects as function of density threshold) and it is therefore very instructive to compare the Betti numbers of two distributions to appreciate how similar or distinct they may be from a topological point of view. For that reason, we plotted on figure 15(b), from top to bottom, the value of β_2 , β_1 , β_0 and the Euler characteristic χ (a topological invariant, computed as the alternate sum of the Betti numbers) as measured in S^{128} and $S_N^{2 \times 128}$ (left and right column respectively). As noted in Sousbie (2010), the notions of persistence pairs and Betti numbers are intimately related: the Betti numbers were readily computed from the persistence pairs, the positive critical point of order $k + 1$ increasing β_k when it enters the excursion and the negative critical point of order k decreasing β_k . Distribution $S_N^{2 \times 128}$ was obtained by adding an equal number of randomly distributed particles to the particles in the N-body simulations S^{128} , and the Betti numbers of the two distributions should therefore give some insight on how topology is affected by Poisson noise. Note that the presence of Poisson noise in $S_N^{2 \times 128}$ affects the PDF of the sampled density by slightly downscaling it (numerous random particles land in large scale void regions, increasing their densities, while few of them affect the high density regions, therefore lower their density contrast, see black plain curves on figure 15). When comparing Betti numbers in the two distributions, one would rather want to know whether the same structures (*i.e.* void, tunnel, component) exist in both distributions though, even if it exists at slightly different densities. It is therefore more important to compare the general shape and amplitude of the Betti number in both distributions than their value at a precise density threshold. Inspecting figure 15(b), it is clear that random particles mainly affect the topological properties of the field around the average density ρ_0 , each Betti number differing of about one order of magnitude in S^{128} (left) and $S_N^{2 \times 128}$ (right) at a level around $\rho/\rho_0 = 1$. The situation largely improves after the cancellation of the lower persistence pairs though and it is striking how the shape and amplitude of the Betti numbers at a level of persistence ratio of $3 \sim 4 - \sigma$ become similar. Note also that β_0 is the Betti number that is the least affected by Poisson noise, and for persistence higher than $3 - \sigma$, the values are almost identical in S^{128} and $S_N^{2 \times 128}$. This means that individual components in the Filtration

are created and merge in a very similar way independently of the presence of Poisson noise, which does not affect the filamentary structure of S^{128} . It is therefore reasonable to trust the filaments detected at persistence levels higher than $\sim 3 - \sigma$ as being true topological properties of the underlying distribution. One should nonetheless remain cautious with the identification of voids and wall. In fact, although the topology of the 1-cycles and 2-cycles seems to be correctly recovered in $S_N^{2 \times 128}$ at a significance level of $3 \sim 4 - \sigma$, this is not the case anymore at higher levels and one should be careful not to set the threshold too high. In fact, the cosmological voids and walls are more affected by Poisson noise as they usually live at densities around $\rho/\rho_0 = 1$ where the influence of Poisson noise is maximal and the corresponding persistence pairs have statistically lower persistence ratios than that associated to filaments.

5 CONCLUSION

We implemented DisPerSE (Sousbie 2010) on realistic 3D dark matter cosmological simulations and observed redshift catalogues from the SDSS DR7. We showed that DisPerSE traces very well the observed filaments, walls, and voids seen both in simulations and observations. In either setting, filaments are shown to connect onto halos, outskirt walls, which circumvent voids, as is topologically required by Morse theory. Indeed, DisPerSE warrants that all the well-known and extensively studied mathematical properties of Morse theory are ensured by construction at the mesh level. As illustrated in sections 3, DisPerSE assumes nothing about the geometry of the survey or its homogeneity, and yields a natural (topologically motivated) self-consistent criterion for selecting the significance level of the identified structures. We demonstrated that the extraction is possible even for very sparsely sampled point processes, as a function of the persistence ratio (a measure of the significance of topological connections between critical points), which allows us to account consistently for the shot noise of real surveys. The corresponding recovered cosmic web is also “persistent” in as much as it is robust because it relies on intrinsic topological features of the underlying density field. Hence we can now trace precisely the locus of filaments, walls and voids from such samples and assess the confidence of the post-processed sets as a function of this threshold, which can be expressed relative to the expected amplitude of shot noise. DisPerSE also seems to be robust, in as much that more sparsely samples recover filamentary structures which are consistent with those of the better sampled catalogues. In a cosmic framework, this criterion was shown to level with FoF structure finder for the identifications of peaks, while DisPerSE also identifies the connected filaments and quantitatively produces on the fly the full set of Betti numbers (number of holes, tunnels, connected components etc...) *directly from the particles*, as a function of the persistence threshold, as these follow from the persistence pairs. We investigated the evolution of the critical points, the Betti numbers and the Euler characteristic as a function of the persistence ratio: it illustrates the biases involved in filtering low persistence ratios. For dark matter simulations, this criterion was shown to be sufficient

even if one particle out of two is noise, when the persistence ratio is set to $3\text{-}\sigma$ or more. We applied this procedure to the localization of a specific filamentary configuration and observed an “optically faint” cluster at a galaxy filaments junction, identified in the SDSS catalogue. An X-ray counterpart could indeed be observed (Kawahara et al. in prep) by the X-ray satellite SUZAKU. The filaments of the SDSS extracted with DisPerSE are available online at the URL <http://www.iap.fr/users/sousbie/SDSS-skeleton.html> as a set of segments with extremities in RA, DEC, redshift. All these results are very encouraging for future investigations using DisPerSE, for searching galaxy clusters, galaxy groups, and missing baryons of the universe in particular, and for the study of LSS in general.

Acknowledgements

The authors thank T. Nishimichi for his help dealing with SDSS galaxy catalogue, H. Yoshitake for his help with X-ray analysis and TS thanks Y. Suto for his constant help and support.

This work was made possible through an extensive usage of the Yorick programming language by D. Munro (available at <http://yorick.sourceforge.net/>) and also CGAL, the Computational Geometry Algorithms Library, (<http://www.cgal.org>), to compute the Delaunay tessellations.

The filaments of the SDSS extracted with DisPerSE is readily available online at the URL <http://www.iap.fr/users/sousbie/SDSS-skeleton.html>.

Funding for the SDSS has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>. TS gratefully acknowledges support from JSPS (Japan Society for the Promotion of Science) Postdoctoral Fellowship for Foreign Researchers award P08324. HK is supported by a JSPS (Japan Society for Promotion of Science) Grant-in-Aid for science fellows. CP acknowledges supports from a Leverhulme visiting professorship at the Astrophysics department of the University of Oxford. This work is also supported by Grant-in-Aid for Scientific research from JSPS and from the Japanese Ministry of Education, Culture, Sports, Science and Technology (No. 22-5467).

REFERENCES

- Abazajian K. N., et al., 2009, *ApJ Sup.*, 182, 543
 Aracil B., Petitjean P., Pichon C., Bergeron J., 2004, *A&A*, 419, 811
 Aragon-Calvo M. A., Platen E., van de Weygaert R., Szalay A. S., 2008, ArXiv e-prints
 Aragon-Calvo M. A., van de Weygaert R., Araya-Melo P. A., Platen E., Szalay A. S., 2010, *MNRAS*, 404, L89
 Blanton M. R., Schlegel D. J., Strauss M. A., Brinkmann J., Finkbeiner D., Fukugita M., Gunn J. E., Hogg D. W., Ivezić Ž., Knapp G. R., Lupton R. H., Munn J. A., Schneider D. P., Tegmark M., Zehavi I., 2005, *AJ*, 129, 2562
 Blumenthal G. R., da Costa L. N., Goldwirth D. S., Lecar M., Piran T., 1992, *ApJ*, 388, 234
 Buote D. A., Zappacosta L., Fang T., Humphrey P. J., Gastaldello F., Tagliaferri G., 2009, *ApJ*, 695, 1351
 Cen R., Ostriker J. P., 1999, *ApJ*, 514, 1
 Cohen-Steiner D., Edelsbrunner H., Harer J., 2007, *Discrete Comput. Geom.*, 37, 103
 Danforth C. W., Stocke J. T., Shull J. M., 2010, *ApJ*, 710, 613
 de Lapparent V., Geller M. J., Huchra J. P., 1986, *ApJ Let.*, 302, L1
 Edelsbrunner H., Letscher D., Zomorodian A., 2000, in , 41st Annual Symposium on Foundations of Computer Science (Redondo Beach, CA, 2000). IEEE Comput. Soc. Press, Los Alamitos, CA, pp 454–463
 Edelsbrunner H., Letscher D., Zomorodian A., 2002, *Discrete Comput. Geom.*, 28, 511
 Efstathiou G., Davis M., White S. D. M., Frenk C. S., 1985, *ApJ Sup.*, 57, 241
 Fang T., Bryan G. L., Canizares C. R., 2002, *ApJ*, 564, 604
 Fang T., Buote D. A., Humphrey P. J., Canizares C. R., Zappacosta L., Maiolino R., Tagliaferri G., Gastaldello F., 2010, *ApJ*, 714, 1715
 Forero-Romero J. E., Hoffman Y., Gottlöber S., Klypin A., Yepes G., 2009, *MNRAS*, 396, 1815
 Forman R., 2002, *Sém. Lothar. Combin.*, 48, Art. B48c, 35 pp. (electronic)
 Fukugita M., Hogan C. J., Peebles P. J. E., 1998, *ApJ*, 503, 518
 Gay C., Pichon C., Le Borgne D., Teyssier R., Sousbie T., Devriendt J., 2010, *MNRAS*, 404, 1801
 Hahn O., Porciani C., Carollo C. M., Dekel A., 2007, *MNRAS*, 375, 489
 Jost J., 2008, *Riemannian geometry and geometric analysis*, fifth edn. Universitext, Springer-Verlag, Berlin
 Kawahara H., Yoshikawa K., Sasaki S., Suto Y., Kawai N., Mitsuda K., Ohashi T., Yamasaki N. Y., 2006, *PASJ*, 58, 657
 Milnor J., 1963, *Morse theory*. Based on lecture notes by M. Spivak and R. Wells. Annals of Mathematics Studies, No. 51, Princeton University Press, Princeton, N.J.
 Mitsuda K., Bautz M., Inoue H., Kelley R. L., Koyama K., Kunieda H., Makishima K., Ogawara Y., Petre R., Takahashi T., Tsunemi H., White N. E. e. a., 2007, *PASJ*, 59, 1
 Neyrinck M. C., 2008, *MNRAS*, 386, 2101
 Novikov D., Colombi S., Doré O., 2006, *MNRAS*, 366, 1201
 Ohashi T., Ishida M., Sasaki S., Ishisaki Y., Mitsuda K., Yamasaki N. Y., Fujimoto R., Takei Y., Tawara Y., Furuzawa A., Suto Y., Yoshikawa Y., Kawahara H., Kawai N., Tsuru T. G., Matsushita K., Kitayama T., 2006, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 6266 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, DIOS: the diffuse intergalactic oxygen surveyor
 Platen E., van de Weygaert R., Jones B. J. T., 2007, *MNRAS*, 380, 551
 Pogossyan D., Bond J. R., Kofman L., Wadsley J., 1996, in *Bulletin of the American Astronomical Society Vol. 28 of Bulletin of the American Astronomical Society*, The Cosmic Web and Filaments in Cluster Patches. pp 1289–

+

- Pogosyan D., Pichon C., Gay C., Prunet S., Cardoso J. F.,
Sousbie T., Colombi S., 2009, *MNRAS*, 396, 635
- Schaap W. E., van de Weygaert R., 2000, *A&A*, 363, L29
- Sheth R. K., van de Weygaert R., 2004, *MNRAS*, 350, 517
- Sousbie T., Colombi S., Pichon C., 2009, *MNRAS*, 393, 457
- Sousbie T., Pichon C., Colombi S., Novikov D., Pogosyan
D., 2008, *MNRAS*, 383, 1655
- Sousbie T., Pichon C., Courtois H., Colombi S., Novikov
D., 2008, *ApJ Let.*, 672, L1
- Stoica R. S., Martínez V. J., Mateu J., Saar E., 2005, *A&A*,
434, 423
- Stoica R. S., Martínez V. J., Saar E., 2007, Journal of the
Royal Statistical Society: Series C (Applied Statistics) 56
(4), 459-477, 56, 1
- Stoica R. S., Martínez V. J., Saar E., 2010, *A&A*, 510,
A38+
- Summers F. J., Davis M., Evrard A. E., 1995, *ApJ*, 454, 1
- Tago E., Saar E., Tempel E., Einasto J., Einasto M., Nurmi
P., Heinämäki P., 2010, *A&A*, 514, A102+
- Tripp T. M., Savage B. D., Jenkins E. B., 2000, *ApJ Let.*,
534, L1
- van de Weygaert R., 2002, ArXiv Astrophysics e-prints
- Yoshikawa K., Yamasaki N. Y., Suto Y., Ohashi T., Mit-
suda K., Tawara Y., Furuzawa A., 2003, *PASJ*, 55, 879

TERMINOLOGY

Arc An arc is a 1-cell: an integral line (or a V-path in the discrete theory) whose origin and destinations are critical points. The arcs of Morse-Smale complex connect two critical points of order difference 1 (*i.e.* in 2D, a minimum and a saddle-point or a maximum and a saddle-point).

n -cell A n -cell is a region of space of dimension n such that all the integral lines in the n -cell have a common origin and destination. The n -cells basically partition space into regions of uniform gradient flow

Coface A coface of a k -simplex α_k is any p -simplex β_p , with $p \geq k$, such that α_k is a face of β_p . In 3D, the cofaces of a segment (*i.e.* a 1-simplex) are any triangle or tetrahedron (*i.e.* 2 or 3-simplex) whose set of summits (*i.e.* vertexes) contains the two vertexes at the extremities of the segment, as well as the segment itself.

Cofacet A cofacet of a k -simplex α_k is a coface β_{k+1} of α_k with dimension $k + 1$. Equivalently, α_k is a facet of β_{k+1} .

Critical point of order k For a smooth function f , a critical point of order k is a point such that the gradient of f is null and the Hessian (matrix of second derivatives) has exactly k negative eigenvalues. in 2D, a minimum, saddle point and maximum are critical points of order 0, 1 and 2 respectively.

Critical k -simplex A critical k -simplex is the equivalent in discrete Morse theory of the critical point of order k in its smooth counterpart. Note that in 2D, the equivalent of a minimum is a critical vertex (0-simplex), a saddle-point is a critical segment (1-simplex) and a maximum is a critical triangle (2-simplex).

Crystal A crystal is a 3-cell: a 3D region delimited by 6 quads and 12 arcs, within which all the integral lines (or V-paths in the discrete case) have identical origin and destinations.

k -cycle A k -cycle in a simplicial complex corresponds to a k dimensionnal topological feature. in 3D, 0-cycles correspond to independant components, 1-cycles to loops and 2-cycles to shells

Discrete Gradient A discrete gradient of a discrete Morse-Smale function f defined over a simplicial complex K pairs simplexes of K . Within a gradient pair, the simplex with lower value is called the tail and the other the head, and any unpaired simplex is critical.

Discrete Morse-Smale complex (DMC) The discrete Morse-Smale complex (DMC for short) is the equivalent of the Morse-Smale complex applied to simplicial complexes.

Discrete Morse-Smale function A discrete Morse-Smale function f defined over a simplicial complex K associates a real value $f(\sigma_k)$ to each simplex $\sigma_k \in K$.

Excursion set see sub-level set.

Face A face of a k -simplex α_k is any p -simplex β_p with $p \leq k$, such that all vertexes of β_p are also vertexes of α_k . In 3D, the faces of a 3-simplex (*i.e.* a tetrahedron) are the tetrahedron itself, the 4 triangles that form its boundaries, the 6 segments that form its edges, and its 4 summits (*i.e.* vertexes).

Facet A facet of a k -simplex α_k is a face β_{k-1} of α_k with dimension $k - 1$. The facets of a 3-simplex (*i.e.* a tetrahedron) are the 4 triangles (*i.e.* 2-simplexes) that form its boundaries.

Filtration A filtration of a simplicial complex K is a *growing* sequence of sub-complexes K_i of K , such that each K_i is also a simplicial complex. If the different K_i are defined by a discrete function F_ρ as the set of simplexes of K with value $F_\rho(\sigma)$ less or equal to a given threshold, a filtration can be thought of as the discrete equivalent of a sequence of growing sub-level sets of a smooth function.

Gradient pair / arrow A Gradient pair or arrow is a set of two simplex, one being the facet of the other, and such that they are paired within a discrete gradient. Within a gradient pair, the simplex with lower value is called the tail and the other the head.

Integral line An integral line of a scalar function $\rho(\mathbf{x})$ is a curve whose tangent vector agrees with the gradient of $\rho(\mathbf{x})$.

Level set / Sub-level set A level set, also called iso-contour, of a function $\rho(\mathbf{x})$ at level ρ_0 is the set of points such that $\rho(\mathbf{x}) = \rho_0$. The corresponding Sub-level set is the set of points such that $\rho(\mathbf{x}) \geq \rho_0$

Ascending/Descending p -manifold Within a space of dimension d , an ascending p -manifold is the set of points from which, following minus the gradient, one reaches a given critical point of order $d - p$. A descending p -manifold is the set of points from which, following the gradient, one reaches a given critical point of order p . For instance, ascending 1-manifolds in 3D can be associated to the filaments, and ascending 3-manifolds describe the voids

Morse function A Morse function is a continuous, twice differentiable smooth function whose critical points are non degenerate. In particular the eigenvalues of the Hessian matrix (*i.e.* the matrix of the second derivatives) must be non-null

Morse complex The Morse complex of a Morse function is the set of its ascending (or descending) manifolds.

Morse-Smale function A Morse-Smale function is a Morse function whose ascending and descending manifolds intersect *transversely*. This means that there exist no point where an ascending and a descending manifold may be tangent

Morse-Smale complex The Morse-Smale complex is the intersection of the ascending and descending manifolds of a Morse-Smale function. One can think of the Morse-Smale complex as a network of critical points connected by n -cells, defining a notion of hierarchy and neighborhood among them. In particular, the geometry of the arcs (*i.e.* 1-cells) is determined by the critical integral lines (*i.e.* integral lines that join critical points) and the order of two critical points connected by an arc may only differ by 1.

Peak/Void patch In 3D, a peak patch is a descending 3-manifold (*i.e.* the region of space from which, following the gradient, one reaches a given maximum), and a void patch an ascending 3-manifold (*i.e.* the region of space from which, following minus the gradient, one reaches a given minimum).

Persistence The persistence of a persistence pair (or equivalently of the corresponding k -cycle it creates and destroys) is

defined as the difference between the value of the two critical points (or critical simplexes in the discrete case) in the pair. It basically represents its life time within the evolving sub-level sets (or filtration in the discrete case).

Persistence pair In the smooth context of a function ρ , persistence pairs critical points P_a and P_b of ρ that respectively create and destroy a topological feature (or k -cycle) in the sub-level sets of ρ , at levels $\rho(P_a)$ and $\rho(P_b)$. In the discrete case of a simplicial complex K , a persistence pair is a pair of critical simplexes σ_a and σ_b of a given discrete function $F_\rho(\sigma)$, such that σ_a creates a k -cycle (*i.e.* topological feature) when it enters the filtration of K according to F_ρ and σ_b destroys it when it enters.

Persistence ratio The persistence ratio of a persistence pair (or equivalently of the corresponding k -cycle it creates and destroys) is the ratio of the value of the two critical points (or critical simplexes in the discrete case) in the pair. Persistence ratio is preferred to regular persistence in the case of strictly positive functions such as the density field of matter on large scales in the universe.

Quad A quad is a 2-cell : a 2D region delimited by four arcs within which all the integral lines (or V-pathes in the discrete case) have identical origin and destinations.

k -simplex A k -simplex is the k dimensional analog of a triangle: the simplest geometrical object with $k + 1$ summits, called vertex. It is the building block of simplicial complexes

Simplicial complex A simplicial complex K is a set of simplexes such that if a k -simplex α_k belongs to K , then all its faces also belong to K . Moreover, the intersection of two simplexes in K must be a simplex that also belongs to K

Vertex A vertex is a 0-simplex or simply a point.

V-path A V-path is the discrete equivalent of an integral line: it is a set of simplexes linked by discrete gradient arrows and face/coface relation. Tracing a V-path consists in intuitively following the direction of the gradient pairs of a discrete gradient from a critical simplex to another.